NOV 1 9 1929

PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS

HAROLD RUGG, *Chairman.*

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — L. M. TERMAN
FRANK N. FREEMAN — V. A. C. HENMON — BEARDSLEY RUML — E. L. THORNDIKE

VOL. XX
No. 8

NOVEMBER, 1929

$5.00 a year.
70c per copy.

CONTENTS

RESPONSE TO T. KELLEY. *C. Spearman*. 561

THE MENTAL GROWTH CURVE FOR THE BINET TESTS. *L. L. Thurstone and Luton Ackerson* 569

FUNDAMENTAL CONSIDERATIONS IN THE STUDY OF MUSICAL ABILITY. *H. Lowery*. 584

COMPUTATION OF CORRELATION COEFFICIENTS. *Edward E. Cureton*. 588

EXAMINATION AS AN AID TO LEARNING. *Arthur T. Jersild* . . 602

A STUDY OF THE EFFECT OF A DEFINITE COURSE OF READING IN GENERAL LITERATURE UPON ACHIEVEMENT IN CONTENT SUBJECTS WITH CHILDREN OF SUPERIOR MENTAL ABILITY *Cora Lee Danielson*. 610

SYSTEMATIC DIFFERENCES IN THE VARIOUS PARTS OF THE HERRING REVISION OF THE BINET-SIMON TEST WHEN APPLIED TO NORMAL DULL ADULTS. *H. H. Remmers*. 622

THE RELATION OF LEFT-HANDEDNESS TO PSYCHO-NEUROTIC TRAITS. *R. G. Wetmore and G. H. Estabrooks* 628

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES. 630

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION. 636

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street — Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

The Journal of Educational Psychology

is published monthly, by Warwick & York, Inc., Baltimore, Md., and York, Pa. Title page and index are bound in December number.

Manuscripts for publication, books or other materials for review, and news items should be addressed to Harold Rugg, Chairman of the Editorial Board, 600 W. 118th Street, New York City.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single issues are 60 cents each, and less than a full year's subscription is charged at single issue rate for every month ordered.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which change is to take effect. No claim for nonreceipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Warwick & York, Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European
Plan

Entirely
Fire-Proof

Centrally
Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

TEACHING IN COLLEGE AND UNIVERSITY

A Survey of the Problems and Literature in Higher Education

BY CARTER V. GOOD

This book does essentially the same thing in college teaching that the author's *How To Do Research In Education* accomplished in the field of graduate study and research. The book is an exhaustive survey of the experimentation and investigation of the past decade in higher education together with a statement of the attempts made toward solution of pressing problems in this field. The problems considered confront the teacher of any college subject, and the discussion, therefore, will prove of service in classes organized for the training of college, normal school, university, and junior college teachers.

Instructors who wish to consider this text for classroom use this fall are invited to communicate with the publishers.

Price $3.00 + 14¢ post

WARWICK AND YORK, INC.
PUBLISHERS
BALTIMORE MARYLAND

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XX November, 1929 Number 8

RESPONSE TO T. KELLEY

C. SPEARMAN

University of London, University College

The publication by Professor T. Kelley of his "Crossways in the Mind of Man" is beyond doubt an outstanding event in recent psychology. For twenty-five years there has been an uninterrupted output of strenuous and highly controversial research aiming at the analysis of mental abilities into their constituent factors. But throughout this work Kelley, peculiarly well equipped though he is for it, has tantalized and troubled us by keeping almost wholly aloof. Now at last, however, breaking his long silence and discarding his inaction, he has plunged into the middle of the fray. And since he honors me with frequent mention, I will gladly try to second his endeavors with the following response.

I. *Outline of the Main Theory of "Two Factors."*—The whole business has pivoted round what has been called the theory of "two factors." And this itself has fallen into two parts; a *main* or general theory; and, on the other hand, various interpretative "sub-theories." To distinguish the former from the latter is of prime necessity, if we would avoid hopeless confusion.

Briefly stated the main, or general, theory is that every measurement of every ability can be resolved into two factors, of which the one is general, whereas the other is specific. Otherwise expressed, all abilities involve one and the same factor "*g*," whereas every different ability has its own particular factor "*s*." Of course, the case may arise where abilities are only different *in part;* they are then said to "overlap," or to contain a "group factor." But, in general, this only occurs when the abilities are "very closely and obviously allied." We may add that the *g* appears to be functionally indivisible, whereas the *s* can usually be split up with advantage into several

components. The evidence for all this theory was first offered in 1904.[1] In a very large number of subsequent publications, it has continually been rendered more precise, more complete, and more secure from possible objections.

How, then, is this main or general theory now envisaged by Kelley? His initial attitude towards it appears to be non-committal, for early in his book he remarks that he will not attempt either its proof or disproof. His reason given is that, even if he could show that the *s*'s were not specific in any particular case, the defender of the theory could still always elude the attack by saying that in this case the abilities were "very closely and obviously allied" (see above). But on this point, Kelley is hard to follow. No more than other scientists can psychologists really get away with merely "saying" things. They are obliged to bring forward evidence.

Passing on, however, from this initial attitude of Kelley to his actual findings later on in the book, these certainly corroborate the main theory of two factors in its general outline. For he, just like ourselves before him, finds that every ability involves one and the same common factor which may be called *g*. Like us, he finds that every ability has also a factor peculiar to itself. And finally he, no less than we, has furthermore come upon "group factors." In a word, his whole system of results is precisely that which constitutes the essence of the main or general theory of two factors; and which, it may be added, would appear to be totally unlike any other theory of ability ever published.

II. *Kelley's Detailed Findings.*—Indeed, the corroboration really brought by Kelley's work goes far beyond the general outline. It extends to nearly all the major experimental details which he communicates. These almost entirely concern the aforesaid "group factors." Throughout all the years that these have been heatedly disputed, the only school—as far as I know—which has actually investigated them is our own. But now Kelley does so likewise, and with his wonted thoroughness.

Comparing his chief results with ours, he himself writes as follows:

> On the whole the two sets of findings are quite remarkably in harmony, the agreements being in the matter of a spatial, a numerical, a memory, and even a general factor, though this last is differently interpreted, and also in the conclusion that a large number of specific motor (probably also sensory) factors exist. There is scarcely a disagreement in the matters of music, purpose, cleverness, and sex,

[1] "General Intelligence" Objectively Determined and Measured. *American Journal of Psychology*, 1904.

though here the data are inadequate. There does seem to be a real disagreement in the importance and extent of a verbal factor and in that of a mental speed factor.[1]

Surely, a fuller corroboration of the general theory of two factors could hardly be desired. But really the concord is better still. For as regards the "speed factor," Kelley seems not to have understood me; his findings and mine do not really present any conflict (see The Abilities of Man," Chap. XIV). And even as regards the "verbal factor" the alleged discrepancy between us is chiefly due to his taking this term in a very different sense from mine. Moreover I am willing to concede that in the last year or so my own views have received more development on this point than on any other. Under the conditions that led to the conclusions reached by me previously, I still believe that the verbal factor (in my sense of the word) has only a small influence. But under *other conditions*, as our later work has shown, the importance of this factor may, I now think, become very great.

So much for Kelley's major details. Although obtained in a different way, their agreement with the results previously obtained by myself and collaborators is most remarkable. But turning next to the minor details obtained by him, these certainly contain many things which are not to be found in our work. If these novel contributions should eventually be verified, so much the greater gain for science. But, as we shall see later, just these novel results would appear to be of dubious validity.

III. *Interpretation of "g."*—In the two preceding sections—both the outline and the details of the theory of g—we have throughout remained within the confines of psychological observation and mathematical demonstration. But in the present section we pass beyond these; from observation and demonstration we go on to speculative interpretation. Not content with establishing the existence of a value g, we advance into the "sub-theories" which attempt to portray its significance.

Now, as regards these interpretative sub-theories, Kelley and I do seem to disagree. And that this should happen would scarcely be surprising, since the matter is still so very speculative. But the real difference between us is, I think, quite other than that which he takes it to be. On my side, I had suggested that g may measure something of the nature of a psycho-physical energy. He would oppose this by contending that it may arise merely from some heterogeneity in the

[1] "Crossways in the Mind of Men." P. 23.

population tested. But I maintain that these two views are not really opposed to one another; indeed I contest his whole conception of the influence of heterogeneity upon correlations.

Suppose that a population consists of two heterogeneous sections, A and B, and that the difference between them really has produced the value *g*. This can, in general, only happen when A and B differ in some property measured by *g*; and this property might just as well consist in an energy as in anything else. But why then, it may be asked, does one often take such trouble to eliminate some heterogeneity from the population? Only in order to show, I reply, that the *g* (or other result at issue) has not been caused by this kind of heterogeneity *alone*, but also by other kinds. Eliminate heterogeneity of all kinds whatever, and you will, at the same time, have eliminated all correlation!

IV. *Recriminations.*—From the experimental results let us pass on to the question of technique. On this head Kelley makes certain charges against myself, and with an emphasis and repetition that are quite disconcerting.

Thus, he charges me over and over again with having "almost entirely neglected such factors as those of sex, age, and race."[1] He writes that I appear "really to have given very little thought to" them.[2] I am said "to have been guilty of serious oversight about them." He urges that

> Spearman's groups typically have not been children of the same age, and he has not resorted to a partial correlation technique to reduce his data to a constant age basis.[3]

On reading all this, I had to wonder whether I was really awake! Why, already in 1904—years and years before any one else ever troubled themselves with such factors—I made them the chief topic of two successive papers, myself introducing this very "partial correlation technique" of which he now talks.[4] Two years later, I wrote a paper conjointly with Krueger in which these same factors were set forth with even greater emphasis. Ever since then, my collaborators, my students, and I myself have been continually both

---

[1] *Ibid.*, p. 210.

[2] *Ibid.*, p., 212.

[3] *Ibid.*

[4] See footnote on p. 563. Also the "Proof and Measurement of Association between Two Things," in the same volume. Also, *Zeitschrift fur Psychologie*, Vol. XLIV, 1906.

urging others to consider these factors and diligently investigating them ourselves.

But now comes the strangest fact of all. Or is it perhaps not so strange? In this very work of Kelley, where he so insistently charges me with failing to partial out such factors as age and sex, and where his own main arguments depend so vitally upon the alleged influence of such factors, he himself from beginning to end totally omits to partial them out or otherwise estimate their influence. Surely here, if anywhere, a Freudian might fairly suggest the mechanism of "projection."

Nor does this amusing case stand isolated. For he furthermore accuses my technique with involving unknown "probable errors," while in truth the investigation of these has been our main work for many years and our publications are crowded with applications of them. The sole explanation of his charge would seem to be that his own technique really does involve unknown probable errors.[1]

Almost as much may be said about his repeated reproaches that I have always employed "small populations" (apparently he means what are usually named small "samples"). He writes, "Spearman's populations have been very small,"[2] "Spearman's studies have been found wanting . . . because of the small populations involved,"[3] "Spearman has been guilty of . . . serious oversight . . . and that is a failure to appreciate that his point of view cannot be statistically established by a study of small populations."[4] As a matter of fact, my own conclusions have been based on an aggregate of samples amounting to at least 10,000 subjects, whilst his rest upon 357 in all. Turning to the single samples, my largest was 2,599,[5] whereas his is 140. Indeed, he himself quotes several of the researches used by me where the samples were larger than his largest.

Nevertheless, behind all these pleasant allegations there does appear to exist a serious difference of opinion between us which stands in urgent need of settlement. The ground of his objections to my supposedly small samples is their inadequacy to prove what he takes to be my thesis; namely, that (under the stated conditions) no group factors exist. But this has *never* really been my thesis; far from attempting to prove any such thing, I believe it to be absolutely

[1] "Crossways in the Mind of Men." Pp. 15, 84, 119, 127.
[2] *Ibid.*, p. 15.
[3] *Ibid.*, p. 191.
[4] *Ibid.*, p. 214.
[5] "The Abilities of Man." P. 153.

incapable of proof, and certainly not ever provable by increasing the size of the samples. For if no group factor should be manifested even with a thousand subjects, there is still nothing to prevent it from becoming manifest with 10,000; or if even this failed, with a million. All that I have really ever attempted to prove is that (under the said conditions) no group factor has appeared which is significant in the sense of *exceeding the limits of the experimental error*. More than this can never be proved about any negative thesis in any science. The experimental error can indeed be reduced in size, but never abolished. All actual results are at best but approximations to the truth. So too, probably, are all our present theories. Far from taking that of "Two Factors" to be an exception, I expressly stated in the very first statement of it that it was subject to "inevitable eventual corrections and limitations."[1]

There is another point, and one whose practical importance is far greater still, on which Kelley and I would seem to differ; this concerns the usage of the "probable errors." My co-workers and myself hold that the burden of proof of a group factor lies with him who asserts its existence, and the rule of our laboratory (both on the present matter and all others) is that, for complete statistical proof a result ought not to be much less than five times the probable error. This agrees well enough with the rule of the wise Udny Yule, who says that a result, in order to be certainly significant, should be at least three times the standard error (and therefore about four and a half times the probable error).[2] Results only about three times the probable error we take to be suggestive, indeed, for further investigation, but not by themselves conclusive. Very different from all this appears to be the procedure of Kelley. He sharply reproaches me for refusing to accept a result of .20 when its probable error is .13.[3] Naturally enough, then, a result that is twice the probable error appears to him quite satisfactory.[4]

To examine this divergence of view a little closer, let us suppose that, as commonly happens, the number of observed "tetrad differences" comes to a few hundreds. Then the mere errors of sampling *ought* on an average to reduce over a score of cases which are more than twice the probable error; it *ought* to produce some dozen or so over three times the error; and at least a case or two about four times the

[1] *American Journal of Psychology*, 1904.

[2] "Introduction to the Theory of Statistics," 1919.

[3] "Crossways in the Mind of Man." P. 214.

[4] *Ibid.*, p. 215; see also p. 64.

error. Thus if we followed Kelley, we should probably flood the scientific records with results that really have no other basis than that of sampling errors.

Does any such disaster actually befall Kelley himself? Grounds are not lacking to suspect as much. For instance, when he records that a test, specially constructed by him to depend upon speed, is actually correlated with his "speed factor" inversely. Another of these suspicious cases is when his "spatial factor" correlates less with both his expressly constructed tests of this space factor than it does with his test of memory for meaningless syllables. Yet another is when his expressly constructed test of meaningless memory does not correlate at all with his memory factor, but goes inversely with his verbal factor. On employing the probable errors in a more reasonable manner, all such intrinsic absurdities vanish and the upshot is, as Holzinger has so admirably demonstrated,[1] that Kelley's results admit of far simpler expression and actually fit the theory of "Two Factors" even more precisely than mine have usually done.[2]

V. *Essential Differences of Analytical Technique.*—The preceding differences between Kelley and myself are, however, of a very general nature. Much as they may or may not invalidate his results and conclusions, they do not affect the principles specially characteristic of his new technique for factorizing abilities. Let us then for a moment turn to this.

It appears to consist of two parts quite independent of each other. The first of these involves some very interesting mathematics. Setting out from considerations similar to those of Garnett to whom he scarcely allows due credit, he presents us with several theorems for factorizing abilities (or other variables) on the basis of their intercorrelations. Unfortunately, these theorems have not as yet proved of much practical service. He himself writes frankly enough that their usage

> does fail as a desired technique when the number of factors exceed three, and it even is insufficient for two or three factors in that it does not tell just where the various factors are located and what are the numerical values of their standard deviations. On these accounts another method has been developed.[3]

The last sentence is the chief one. Actually, it is his second method that he himself employs throughout. What are its features and advantages?

---

[1] *Journal of Educational Psychology*, 1929, p. 91.

[2] *Ibid.*, p. 96.

[3] "Crossways in the Mind of Man." P. 79.

The claim must be rejected, I think (indeed, it does not seem to be raised by Kelley himself, but only by others on his behalf) that the new technique has opened out the way to any class of facts hitherto inaccessible. On the contrary, we have already had occasion to see that every kind of result obtained by him could be compared with a corresponding result already obtained by us.

Nor can I admit his own claim that, unlike myself, he studies a large number of variables simultaneously in a single investigation; for actually, his largest number of simultaneous variables seems to have been thirteen; ours, nearly seventy. Where we do seem to have differed from him is that we have aimed more at getting a set of variables which, however numerous in themselves, tended mainly to illumine one single problem at a time. Whereas he seems to have aimed more at setting simultaneously several disconnected problems.

But the fundamental difference between his technique and ours lies, elsewhere, it seems to me. We, for our part, have always *started off* from the simple case where a certain number of abilities are not related by group factors. Working from this basis, the group factors are then introduced, one by one. Kelley, on the other hand, starts off with group factors already present. A minor difference between us is that, although both make use of the method of successive approximations,[1] he has combined this device with the method of "least squares."

Comparing the advantages of these two techniques, I would say that Kelley's is unquestionably much more liable to error. But on the other hand, it does furnish a last resource to be tried when the starting basis needed by the other technique is not available. Be that as it may, we may at any rate for the present rejoice that the two different techniques, though applied to problems still highly controversial, have on all major points led to extraordinarily harmonious results. By such convergent lines of evidence it is that psychology will attain to the desired status of a genuine science.

---

[1] Mental Tests of Dementia. *Journal of Abnormal Psychology*, 1914.

THE MENTAL GROWTH CURVE FOR THE BINET TESTS*

L. L. THURSTONE

University of Chicago,

AND

LUTON ACKERSON

Institute for Juvenile Research

Until recently it has not been possible to study the nature of the mental growth curve because we have had no unit of measurement for mental growth which represents a constant increment of mentality for the different levels between birth and maturity. The gross scores in mental tests are obviously not units of measurement in this sense because an increment of one point in gross score from ten to eleven does not necessarily represent the same increment in mentality as the interval of one point from one hundred to one hundred one. Gross scores are clearly out of the question as measures of mental growth except insofar as a rank order may be called measurement.

The mental age has the same defect as the gross score as a measurement of mentality. If we have a mental age test perfectly standardized, the mental year from three to four is never regarded as representing the same increment in mentality as the mental year from twelve to thirteen. Mental ages and gross scores can be used for arranging the subjects in rank order but they cannot be called measurements in any true sense because they are not based on any unit of measurement which can be conceded to represent the same increment of mentality in the different parts of the scale.

For many practical purposes it does not matter whether the description of mental development is in any sense a true measurement. In many situations and for many problems it is sufficient that the scores and mental ages and points and percentile ranks arrange the subjects in rank order. We can then say that one subject is superior to another in the particular function tapped by the particular test. Often this is all that is necessary. But for some theoretical problems these quantitative indices are of no use whatever. It is for the solu-

---

*This is one of a series of studies by members of the Behavior Research Fund staff of the Illinois Institute for Juvenile Research, Series B, No. 127.

tion of the problem of measuring mental growth that the method of absolute scaling was devised.[1]

It is a matter of common observation among psychologists in this field that the increment in mental development seems to be larger in the younger children than in the older children for the same time interval such as one year. It is the common belief that the child advances farther in mental development during the second year than during the tenth year. This common belief is sometimes represented graphically by a hypothetical mental growth curve which is concave upwards with a negative acceleration. The problem of defining this curve more exactly has remained unsolved for lack of a valid unit of measurement. It is the purpose of this study to apply the method of absolute scaling to a comprehensive set of data over a wide age range so as to draw a mental growth curve with an absolute unit of measurement and to investigate its shape. We start with no presupposition about the shape of this curve. The assumption of this method that the distribution of absolute test intelligence is normal in each of the age groups is verified at each step by a test which has been previously described.[1]

In Table I we have a summary of the mental ages of 4208 white children who have been given the Stanford-Binet tests at the Institute for Juvenile Research. The Institute records include cases ranging in age from one to twenty-one years but we did not consider it legitimate to include the whole age range. The youngest children could not be examined by the Stanford-Binet examination. Their mental ages were determined by the Kuhlman tests and by supplementary tests. In order to retain uniformity in the tests used, we started our absolute scaling at three years of age. The upper age groups could not legitimately be included because the Institute does not receive any cases beyond age seventeen except under special circumstances. Consequently the children beyond age seventeen cannot be relied upon to represent the same conditions of selection as the regular Institute cases. We have, therefore, limited ourselves to the age range three to seventeen inclusive in the hope that this population may represent fairly uniform conditions of selection as far as the present problem is concerned.

The data of Table I are here summarized in class intervals of one mental year by the Binet tests. In our calculations we used the original data which were tabulated in class intervals of two, three, four, and six months for the different age groups. Consequently

our scaling was done with many more overlapping points than would be possible with class intervals as coarse as one year.

TABLE I.—FREQUENCY DISTRIBUTIONS OF MENTAL AGE FOR WHITE CHILDREN EXAMINED AT INSTITUTE FOR JUVENILE RESEARCH

Chronological Age

MA	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	10	6	3	2	0	1	0	1	1	..	..	1			
1	12	11	13	10	6	3	2	1	1	..	1	2	1	..	1
2	17	18	22	11	16	11	8	9	2	6	3	1	0	..	3
3	17	20	15	17	14	12	15	10	6	6	4	2	4	4	1
4	8	25	32	23	31	18	17	9	6	10	3	4	1	1	1
5	..	13	45	53	46	33	16	10	14	6	10	9	4	2	2
6	..	2	14	57	58	62	53	35	17	20	13	11	9	7	5
7	..	..	3	13	55	71	65	66	35	23	25	17	10	9	5
8	..	..	1	4	20	44	90	71	62	64	46	37	28	16	8
9	..	..	..	1	5	35	56	73	78	70	90	62	41	23	17
10	..	..	..	..	3	4	21	45	49	74	80	63	56	30	14
11	..	..	..	..	2	0	8	16	37	32	46	53	43	36	19
12	..	..	..	..	..	..	2	6	19	30	39	55	56	32	20
13	..	..	..	..	..	..	..	6	19	22	30	52	37	32	17
14	..	..	..	..	..	..	..	1	6	11	12	46	37	36	13
15	..	..	..	..	..	..	..	..	1	5	9	34	31	17	18
16	..	..	..	..	..	..	..	..	1	1	3	24	14	14	7
17	..	..	..	..	..	..	..	..	..	2	1	10	4	7	3
18	..	..	..	..	..	..	..	..	..	..	1	1	6	5	1
19	..	..	..	..	..	..	..	..	..	..	..	..	1	2	0
20	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1
	64	95	148	191	256	294	353	359	354	382	416	484	383	273	156

Total = 4208 cases

The treatment of these data by the method of absolute scaling gives for each age group its standard deviation in Binet test intelligence and its mean test performance. In our calculations we used the standard deviation of the seventeen-year old children as a unit of measurement for variability, and we used the mean test performance of the four-year children as a tentative arbitrary origin for mean test performance.

The calculations of absolute scaling give the ratio between the standard deviations of adjacent age groups. The ratios are the constants k in observation equations of the following form:

$$\sigma_4 = k_{45} \cdot \sigma_5$$
$$\sigma_4 = k_{46} \cdot \sigma_6$$
$$\sigma_5 = k_{56} \cdot \sigma_6$$
$$\sigma_5 = k_{57} \cdot \sigma_7$$
$$\sigma_6 = k_{67} \cdot \sigma_7$$
$$\sigma_6 = k_{68} \cdot \sigma_8$$

We determined these ratios not only for the adjacent age groups such as 4–5, 5–6, 6–7, 7–8, etc. but also for age groups two years apart such as 4–6, 5–7, 6–8, 7–9, etc. The observation equations in the above form were solved by the method of least squares so as to obtain the best value for the standard deviation of each age group. In these calculations the standard deviation for the seventeen-year

TABLE II

Age	σ	M
3	.122	−.121
4	.1621	.000
5	.2327	.119
6	.2677	.294
7	.3574	.454
8	.4208	.635
9	.5077	.829
10	.5746	1.032
11	.6762	1.318
12	.7241	1.432
13	.6975	1.557
14	.9088	1.984
15	.8676	2.024
16	.9265	2.138
17	1.0000	2.060

group was chosen as a unit of measurement so that all of the standard deviations in Table II refer to this unit. It is entirely arbitrary which age group is chosen for the unit of measurement since the final growth curve is not affected by this choice. We found that the least square solution gives practically the same results as the simpler solution previously described so that a detailed account here of the least square solution is considered superfluous.

When the best fitting values of the standard deviations had been determined by absolute scaling, the mean test performance of each age group was determined. These calculations were similar in form to those which have been previously described. The mean test performance of the four-year old children was chosen as a tentative arbitrary origin, and the standard deviation in test intelligence of the seventeen-year old children was used as a unit of measurement. With this arbitrary origin and with this unit of measurement the mean test performance of each age group could be ascertained and plotted. The mean test performance with this scale is shown in Table II for each age group.

The mean test performance for each age group was determined by equations of the following form in which the standard deviations were known. Let sigma values of the higher age group be plotted as ordinates against the sigma values of the lower age group as abscissas. Then the difference between two mean performances on the absolute scale is

$$M_2 - M_1 = y_{21} \cdot \sigma_2$$

in which y_{12} is the y-intercept of the linear plot. Each equation of the above form gives the increment in Binet test intelligence between the mean performances of two age groups. With the mean test performance of the four-year old children chosen as an arbitrary origin we can draw the mental growth curve. The standard deviations of Binet test intelligence for each age group and the mean test performances are both listed in Table II.

With these data we first draw the diagram of Figure 1. This is a plot of variability against mean test performance. The purpose of this graph is to locate the absolute zero for the mental growth curve for the Binet tests. A straight line is fitted to the data by inspection and it is immediately seen that only a very short extrapolation is needed to ascertain the x-intercept. This intercept is the value of the mean test performance at which variability vanishes and it is the mean test performance which has been previously defined as the absolute zero.[2] This point is located according to the present data at $-.43\sigma_{17}$ below the mean test performance of the four-year old children and the standard deviation of the seventeen-year olds was arbitrarily chosen as a unit of measurement. The location of the absolute zero is in no way affected by the choice of the arbitrary origin and unit of measurement for the calculations. The absolute zero may be located from the mean test performance of any of the age groups as an origin

and with the standard deviation of test intelligence of any age group as a unit of measurement.

It will be seen that the graph of Figure 1 is linear. According to the law of variability of test intelligence,[2] the variability of any age group should be proportional to the absolute mean test performance of that age group. According to this law the present diagram should be linear. Inspection of the diagram shows that it can be fitted with a straight line. In the previous study of the absolute zero[2] the age ranges available were short and it was necessary to extrapolate over a

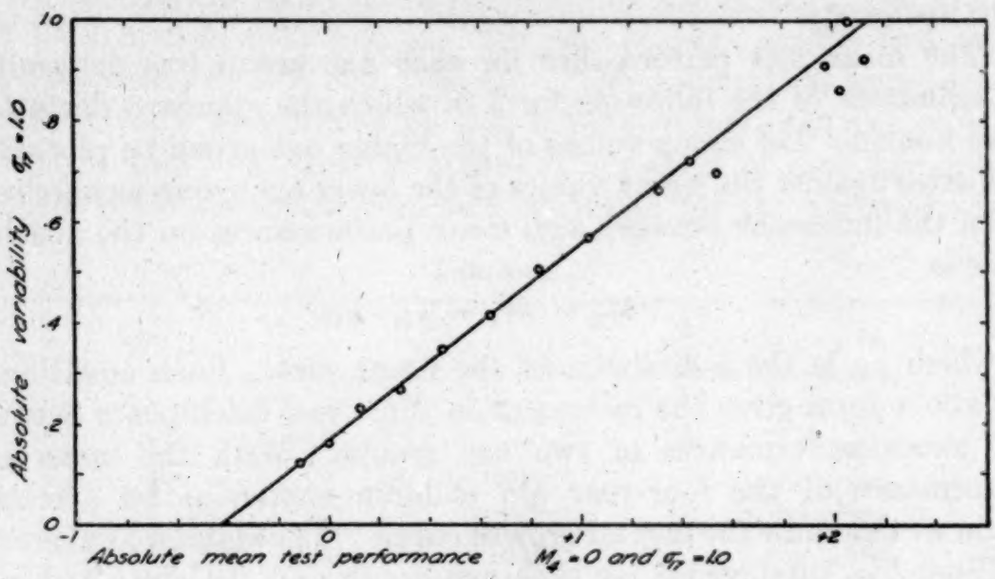

FIG. 1.

considerable age range in order to locate the absolute zero. The present data extend over a much longer age range and the extrapolation is rather short. In Figure 1 we have an excellent verification of the law of variability of test intelligence, according to which this variability is directly proportional to the absolute mean test performance.

The absolute zero has been defined as the mean test performance at which variability vanishes. This value is approximately $-.43\sigma_{17}$ below the mean test performance of the four-year old children. From Table II we may translate this value in terms of σ_4 so that

$$\text{Absolute zero} = M_4 - 0.43\sigma_{17} = M_4 - 2.66\sigma_4 = M_x - 2.66\sigma_x$$

or

$$\frac{\sigma_x}{M_x} = .38$$

In other words, the law of variability of test intelligence can be stated quantitatively for the present population in which the absolute standard deviation of test intelligence is about thirty-eight per cent of the absolute mean test performance for each age. This is undoubtedly higher than would be found in any school population because the Institute handles children of wide variation in mentality including idiots and imbeciles who are examined for committment, children who are examined for major or minor conduct problems, and bright children who are examined for scholarship grants. In any public school population the mean test performance would be higher than in the Institute population but the absolute variability would undoubtedly be smaller. The present data constitute a much more convincing verification of the law of variability of intelligence with age than the data on which the law was first formulated, partly because of the greater age range in the present population which is required to locate the absolute zero.

Our primary interest in this study is the mental growth curve. With the absolute zero located we return to the data of Table II. In Figure 2 we have plotted the mean test performance against chronological age and the absolute zero has also been located on this diagram. The first inspection of this diagram shows that the mental growth curve is a remarkably smooth function for all the age groups up to the age of puberty after which the data become more irregular with a first inversion at the age of seventeen. We had one hundred fifty-six children at age seventeen. Since we have some irregularities in the mental growth curve for the upper ages we cannot be so certain about its upper course but we can discuss with some confidence the general nature of the growth curve for the younger ages up to puberty. We had rather large populations for most of these ages and the growth curve shows remarkable continuity.

Perhaps the principal inference that can be drawn from Figure 2 is that the absolute mental growth curve is positively accelerated in the younger ages. This is contrary to the form of the hypothetical mental growth curve that is rather common in the textbooks. In the light of absolute scaling it seems rather certain that the mental growth curve has an initial positive acceleration and that it is asymptotic to the absolute zero as shown in Figure 2. This conclusion seems more reasonable than the generally assumed negative acceleration because it gives a better picture of continuity for mental growth, regarded as a biological function.

We know with certainty that test intelligence does not continue to rise with age indefinitely. This indicates the presence of an upper asymptote. It is probable that the mental growth curve has some marked deviations from a continuous function at the age of puberty. This makes the mental growth function similar to those of anthropometric measurements which do show irregularities or deviations from a continuous function at the age of puberty. But the presence of such an irregularity at puberty makes it rather difficult to locate the exact course of the mental growth curve of the general population beyond that age. These difficulties we have not attempted to solve

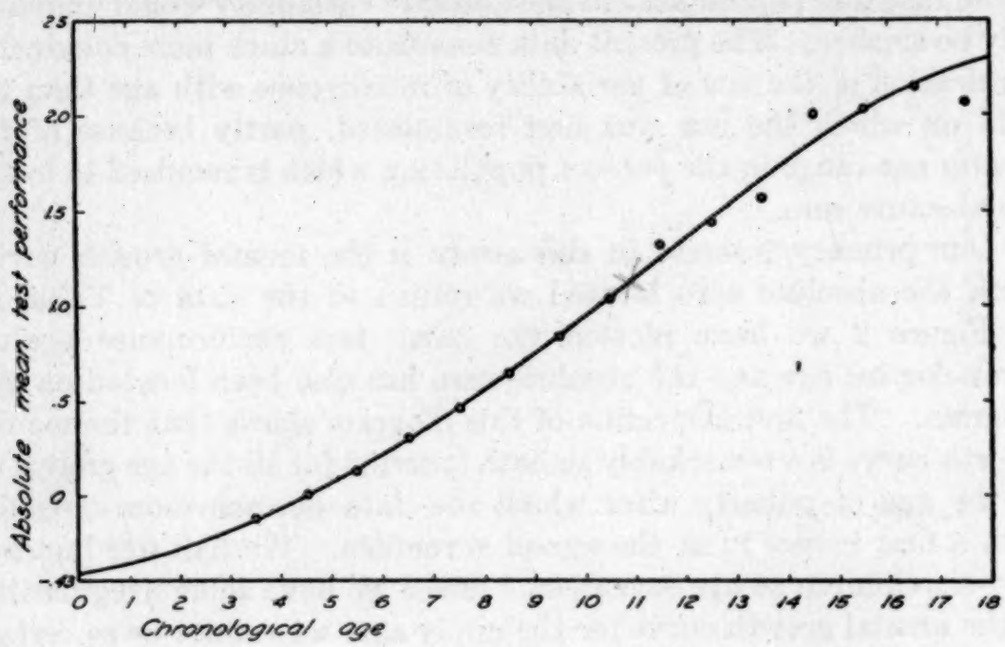

FIG. 2.

with our data. We have merely drawn a smooth curve through the data for the younger ages and we have continued the curve tentatively for the upper ages.

We have no rational procedure for locating the level of Binet test maturity as we have for locating the absolute zero. The growth curve is represented as asymptotic to the level of test maturity. If the curve is asymptotic to absolute zero and also asymptotic to some level of test maturity, then it must have an inflection point. Inspection of our mental growth curve shows that the inflection point is somewhere near the age of eleven.

The population on which this study has been carried out is retarded. In terms of the intelligence quotient the mean performance is about

.8. For this reason the exact ordinates of our curve cannot be taken as representative of the general population but this is not our primary problem. We are concerned with the shape of the mental growth curve. It is likely that the form of the mental growth curve is the same for normal, accelerated, and retarded children but that the constants of the growth curve vary with acceleration and retardation. We assume that the absolute zero is the same for all children. The upper asymptote representing test maturity is probably higher for

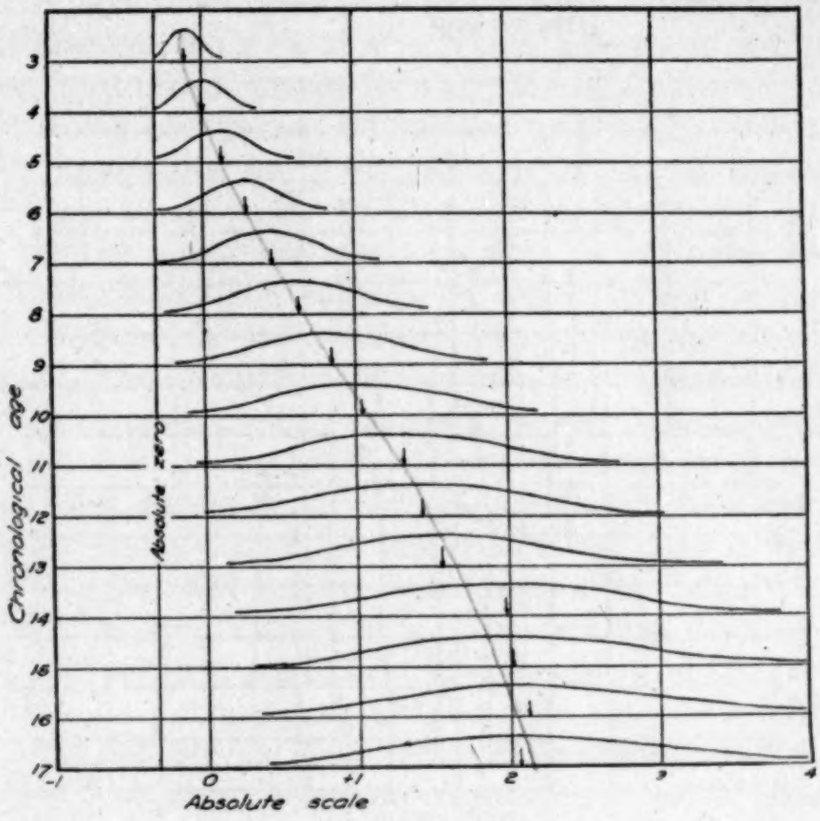

FIG. 3.

bright children than for dull children although this is by no means certain. It is certain, however, that the mental growth curve for bright children approaches the level of the test maturity sooner than that of dull children. It is probable that the mental growth curves for all children have initial positive acceleration, an inflection point, and a later negative acceleration. It is quite probable that bright children reach the inflection point in their growth curves sooner than dull children. A direct comparison between the present data and the absolute scaling of Burt's data on the Binet tests is not feasible because in the present case we have scaled total scores, that is, mental

ages, while the absolute scaling of Burt's data was done on the records of the individual test elements in the Binet scale.

The data of Table II have been reproduced graphically in Figure 3. The base line is the absolute scale with the origin at the mean test performance of the four-year old children in the Institute population. The unit of measurement for this absolute scale is the standard deviation of the seventeen-year old children in the same population. The several age groups have been represented in Figure 3 on separate

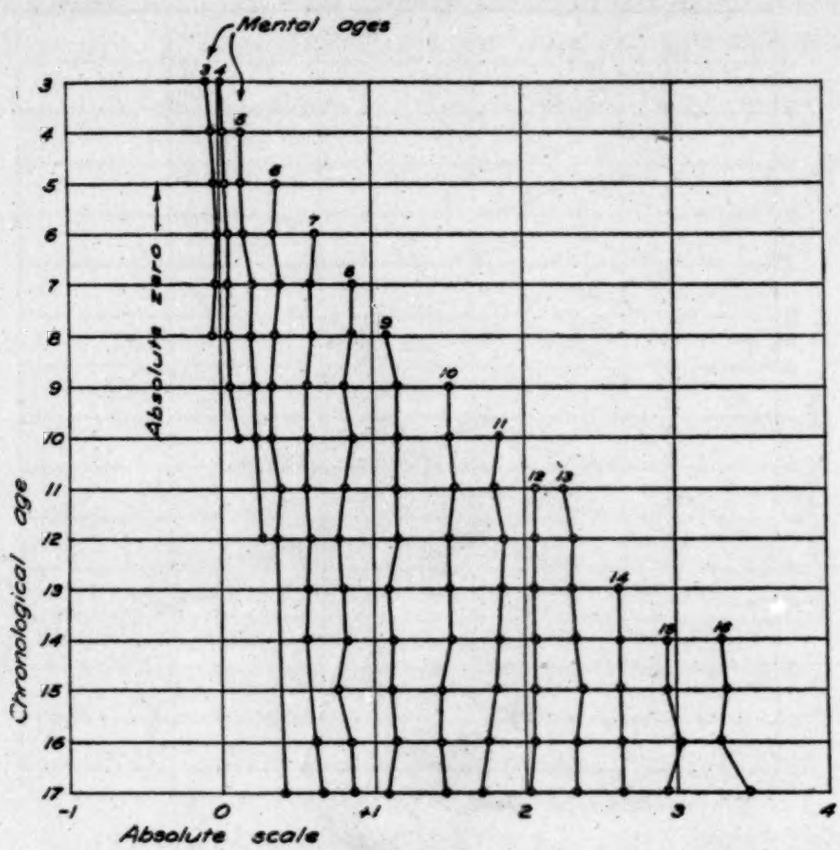

FIG. 4.

base lines in order to avoid confusion and to facilitate comparison. The small arrow heads indicate the mean performances of the several age groups. The dispersions are indicated by the relative spreads of the normal curves. The absolute zero is also shown on this diagram. It is immediately apparent that the variability in absolute test intelligence increases markedly with age. The positive acceleration in the mental growth curve is also shown by the imaginary curve that can be drawn through the arrow heads.

Consider the distribution of seven-year old children in Table I. In that distribution there are thirty-six children with test intelligence

lower than a mental age of four. There are two hundred fifty-six children in this age group so that fourteen per cent of them have intelligence lower than a mental age of four. If we assume, subject to verification, that the distribution of test intelligence in this age group is normal, the sigma value of a mental age of four is $-1.08\sigma_7$. But σ_7 on the absolute scale is $.3574\sigma_{17}$ and hence the mental age of four has an absolute scale location $-.386\sigma_{17}$ below the mean performance of the seven-year old children. But the absolute scale value of the mean performance of the seven-year old children is $+.454$ and hence the absolute scale value of a mental age age of four is $+.068$.

We have proceeded likewise for every one of the entries in Table I. In Figure 4 the results are summarized graphically. The absolute scale value for a mental age of four, as determined from the seven-year age group, was found to be $+.068$ and it is represented by a small circle on the horizontal line marked seven. If our scale is really an absolute scale, then the absolute scale values of any particular test performance should be independent of the age groups on which it is determined. The absolute scale value of a mental age of four should be the same no matter which age group is used for its determination. In Figure 4 it will be seen that all of the determinations of the absolute scale value of a mental age of four are practically the same. This is indicated by the fact that the line connecting them is vertical. If they were not the same this line would shift to the left or to the right or it would be zig-zag in appearance. The fact that the lines of Figure 4 are all vertical with only slight fluctuations, and the further fact that there is no uniform drift either to the right or to the left, is a crucial test of the internal consistency of the absolute scaling method.

From the present investigation we can derive several rather fundamental facts concerning the laws of mental growth.

1. The mental growth curve has a positive acceleration up to the general age level of about ten years.
2. The mental growth curve is asymptotic to an adult level.
3. The mental growth curve has an inflection point somewhere in the age range nine to twelve.
4. The inflection point comes earlier for bright children than for dull children.
5. The mental growth curve is asymptotic to absolute zero.
6. The absolute variability of test intelligence increases with age until adult intelligence is attained. It is directly proportional to the

TABLE III.—FREQUENCY DISTRIBUTION OF MENTAL AGE FOR CHILDREN EXAMINED AT THE INSTITUTE FOR JUVENILE RESEARCH, AGES THREE TO SIX INCLUSIVE

CA		3		4		5		6	
MA		B	G	B	G	B	G	B	G
Yr.	Mo.								
0	0–2	0	3	0	0	0	0	0	0
	3–5	1	0	0	1	0	0	0	0
	6–8	3	1	1	2	0	2	1	1
	9–11	2	0	2	0	0	1	0	0
1	0–2	2	3	2	0	5	1	2	0
	3–5	2	0	0	0	1	1	3	0
	6–8	2	1	3	2	2	3	2	2
	9–11	2	0	3	1	0	0	1	0
2	0–2	1	1	5	4	1	7	1	0
	3–5	3	3	2	0	2	2	0	0
	6–8	5	3	2	2	4	3	8	1
	9–11	1	0	2	1	3	0	0	1
3	0–2	4	3	10	2	2	1	4	1
	3–5	4	2	2	0	4	1	0	1
	6–8	2	0	2	2	3	2	4	3
	9–11	2	0	2	0	0	2	4	0
4	0–2	1	2	5	7	7	1	6	0
	3–5	0	1	3	0	2	2	1	2
	6–8	3	1	5	4	7	4	4	2
	9–11	0	0	1	0	5	4	5	3
5	0–2	..	..	0	6	9	9	9	3
	3–5	..	..	1	1	4	4	8	4
	6–8	..	..	1	3	8	4	11	3
	9–11	..	..	0	1	2	5	11	4
6	0.2	..	..	0	1	4	2	11	11
	3–5	..	..	0	0	1	0	7	1
	6–8	..	..	0	0	4	3	12	8
	9–11	..	..	1	0	0	0	3	4
7	0–2	..	..	..	..	1	1	5	3
	3–5	..	..	..	..	0	0	2	0
	6–8	..	..	..	..	0	1	0	3
	9–11	..	..	..	..	0	0	0	0
8	0–2	..	..	..	..	0	0	0	0
	3–5	..	..	..	..	0	1	1	2
	6–8	..	..	..	..	..	..	1	0
	9–11	..	..	..	..	..	..	0	0
9	0–2	..	..	..	..	..	..	0	1
		64		95		148		191	

TABLE IV.—FREQUENCY DISTRIBUTION OF MENTAL AGE FOR CHILDREN EXAMINED AT THE INSTITUTE FOR JUVENILE RESEARCH, AGES SEVEN TO TEN INCLUSIVE

CA		7		8		9		10	
MA		B	G	B	G	B	G	B	G
Yr.	Mo.								
0	0–3	0	0	0	0	0	0	0	0
	4–7	0	0	1	0	0	0	0	0
	8–11	0	0	0	0	0	0	1	0
1	0–3	0	2	1	1	0	0	0	0
	4–7	1	2	0	0	2	0	0	0
	8–11	1	0	1	0	0	0	1	0
2	0–3	2	1	1	0	3	0	1	0
	4–7	6	1	2	2	3	0	3	1
	8–11	3	3	5	1	2	0	2	2
3	0–3	4	2	2	1	2	1	2	2
	4–7	2	3	4	3	3	6	2	1
	8–11	3	0	1	1	3	0	2	1
4	0–3	7	4	8	3	3	3	2	3
	4–7	8	4	0	0	4	2	1	0
	8–11	2	6	5	2	2	3	2	1
5	0–3	6	4	3	4	2	1	0	0
	4–7	6	5	11	3	3	3	5	1
	8–11	14	11	12	0	4	3	2	2
6	0–3	13	6	13	6	11	8	6	3
	4–7	11	6	11	6	9	5	5	4
	8–11	9	13	18	8	15	5	10	7
7	0–3	11	5	21	8	13	5	15	4
	4–7	15	6	13	8	15	8	12	10
	8–11	11	7	12	9	17	7	14	11
8	0–3	4	4	10	6	25	9	16	6
	4–7	7	2	11	3	18	18	16	9
	8–11	3	0	11	3	12	8	13	11
9	0–3	0	1	12	8	13	14	12	13
	4–7	0	1	5	5	8	9	18	8
	8–11	1	2	4	1	10	2	12	10
10	0–3	1	0	3	0	1	8	14	8
	4–7	0	1	0	0	7	4	8	2
	8–11	1	0	1	0	1	0	8	5
11	0–3	0	1	0	0	3	1	4	2
	4–7	0	1	..	..	0	1	7	0
	8–11	0	0	..	..	1	2	3	0
12	0–3	..	..	..	..	0	0	1	1
	4–7	..	..	..	..	1	0	1	0
	8–11	..	..	..	..	1	0	3	0
13	0–3	..	..	..	..	..	..	1	0
	4–7	..	..	..	..	..	..	2	0
	8–11	..	..	..	..	..	..	3	0
14	0–3	..	..	..	..	..	..	0	0
	4–7	..	..	..	..	..	..	1	0
	8–11	..	..	..	0	..	..	0	0
		256		294		353		359	

TABLE V.—FREQUENCY DISTRIBUTION OF MENTAL AGE FOR CHILDREN EXAMINED AT THE INSTITUTE FOR JUVENILE RESEARCH, AGES ELEVEN TO SEVENTEEN INCLUSIVE

CA		11		12		13		14		15		16		17	
MA		B	G	B	G	B	G	B	G	B	G	B	G	B	G
Yr.	Mo.														
0	0–5	0	0	..	..	..	..	1	0						
	6–11	0	1	..	..	..	..	0	0						
1	0–5	1	0	..	..	..	..	0	2						
	6–11	0	0	0	0	0	1	0	0	0	1	..	..	1	0
2	0–5	0	0	1	2	0	0	1	0	0	0	..	..	1	1
	6–11	1	1	2	1	1	2	0	0	0	0	0	0	0	1
3	0–5	3	1	3	1	1	1	0	1	2	1	2	0	0	0
	6–11	1	1	0	2	0	2	0	1	0	1	1	1	1	0
4	0–5	1	1	2	2	0	2	0	4	0	0	0	0	0	0
	6–11	3	1	5	1	0	1	0	0	1	0	1	0	1	0
5	0–5	4	1	0	0	2	2	2	3	3	0	0	1	0	1
	6–11	8	1	5	1	2	4	2	2	1	0	1	0	0	1
6	0–5	6	1	2	3	6	2	1	3	5	3	3	1	1	2
	6–11	6	4	10	5	3	2	4	3	0	1	2	1	1	1
7	0–5	8	3	5	4	5	5	1	2	3	2	2	1	3	1
	6–11	15	9	8	6	12	3	10	4	4	1	4	2	0	1
8	0–5	20	11	19	9	10	9	10	5	11	3	6	2	3	1
	6–11	23	8	25	11	17	10	12	10	11	3	6	2	4	0
9	0–5	29	16	21	17	34	18	21	11	14	12	5	4	4	4
	6–11	21	12	22	10	23	15	22	8	7	8	5	9	3	6
10	0–5	18	7	24	12	28	14	23	9	14	15	6	11	2	3
	6–11	19	5	25	13	22	16	27	4	14	13	6	7	6	3
11	0–5	18	3	13	3	18	11	15	9	9	11	11	8	6	5
	6–11	14	2	12	4	14	3	18	11	10	13	8	9	3	5
12	0–5	10	6	13	3	7	8	16	13	15	10	11	5	3	8
	6–11	2	1	13	1	17	7	14	12	18	13	11	5	6	3
13	0–5	6	5	9	4	9	9	16	10	13	7	7	9	3	6
	6–11	6	2	9	0	10	2	11	15	8	9	9	7	4	4
14	0–5	1	1	4	3	4	3	17	8	12	11	8	6	3	3
	6–11	3	1	4	0	3	2	12	9	11	3	14	8	4	3
15	0–5	1	0	4	0	3	2	15	5	13	7	8	2	2	4
	6–11	0	0	1	0	3	1	7	7	6	5	3	4	9	3
16	0–5	0	0	1	0	1	0	4	8	4	4	5	5	4	1
	6–11	1	0	0	0	2	0	7	5	6	0	2	2	0	2
17	0–5	..	..	1	0	1	0	7	1	3	1	3	2	0	1
	6–11	..	..	1	0	0	0	2	0	0	0	1	1	2	0
18	0–5	..	..	0	0	1	0	1	0	3	1	3	0	0	0
	6–11	..	..	..	..	..	..	..	..	2	0	1	1	1	0
19	0–5	..	..	..	..	..	..	..	..	..	..	1	0	0	0
	6–11	..	..	..	..	..	..	..	..	0	1	1	0	0	0
20	0–5	..	..	..	..	..	..	..	..	..	..	0	0	0	0
	6–11	..	..	..	..	..	..	..	..	..	..	..	..	0	1
		354		382		416		484		383		273		156	

absolute mean test performance for successive age groups under uniform conditions of selection.

These findings contradict the hypothetical mental growth curve of the textbooks, which is usually shown to have a negative acceleration from birth without an inflection point.

References

1. Thurstone, L. L.: A Method of Scaling Psychological and Educational Tests. *Journal of Educational Psychology*, Vol. XVI, 1925.
2. Thurstone, L. L.: The Absolute Zero in Intelligence Measurement. *Psychological Review*, Vol. XXXV, 1928.

FUNDAMENTAL CONSIDERATIONS IN THE STUDY OF MUSICAL ABILITY

H. LOWERY

Technical College, Huddersfield, England

In view of the great success which has attended the application of intelligence tests in school, it seems quite natural to consider whether corresponding tests may not be found possible in music to enable an estimate to be made of native musical ability. It may be noted that music appears to be, *par excellence*, that subject in which tests of native ability should be successful since the possession of musical gifts is regarded by the majority of people as a special endowment and these gifts are precisely what tests of musical ability aim at locating and estimating.

It is the object of the present paper to emphasize the facts that, in devising a scheme of tests of musical ability, technique and interpretation in music must be distinguished and that the interpretative aspects of musical performance must form the basis of any scheme.

When we consider the factors which constitute the successful performance of, say, an instrumental or vocal solo, we find that they may be resolved into two classes, *viz.*, those which have their origin in the technical skill of the performer and those which clearly depend upon the performer's ability to seize upon the meaning intended by the composer of the music and, through his performance, to reveal this meaning to the auditor.

Now the striking fact about the independence of technique and interpretation in music is this: No individual can achieve success in practical music on the strength of technique alone (since technique is only a means to an end) but even where the conditions are unfavorable for technical development a person may still become a great musician through his insight into the meaning of music, that is, through his ability on the interpretative side. Musical ability is thus not necessarily associated with skill in vocal or instrumental music. Wagner, for example, was but a very indifferent instrumentalist yet he ranks among the world's greatest composers. Again, Schumann, in his anxiety to secure the perfect technique of a virtuoso pianist, invented a piece of apparatus by which he hoped rapidly to gain independence of finger action but his zeal carried him too far in this direction. By overstraining, he injured his hands and so had

to give up all hope of a career as a pianist. Nevertheless, by his teaching and his compositions he will ever be remembered as a great musician.

The extreme divergence between technique and interpretation is perhaps nowhere better shown than in the importance attached nowadays in instrumental music to finger and arm gymnastics away from a musical instrument, as for example in the Macdonald-Smith and similar systems of instrumental playing. It is claimed by these systems that (1) the special exercises given result in a development of certain muscles only employed in playing the instruments in question, (2) the rate of development is much more rapid than by the old-fashioned methods of incessant scale and arpeggio practice and (3) the exercises may all be carried out satisfactorily *away from an instrument.* Just because of the physical aspects of technical development, as emphasized in these exercises, the investigator of musicality is obliged to disregard technique and devote his attention to interpretation which is mental in character.

To exalt in this way the position of interpretation in music at the expense of technique will seem a very inadequate statement of the case to those music teachers who, recognizing the fundamental importance of technique *in actual performance,* have given attention to it to the exclusion of considerations of interpretation. There is ample evidence, however, that music teachers have taken technique to this extreme in view of the mere mechanical displays which are frequently given at concerts. It is only necessary here to refer to two musical criticisms of this attitude though the whole of modern critico-musical literature teems with examples: 1. Mr. Fuller-Maitland in a lecture before the Music Teachers' Association on February 5, 1916[1] after acknowledging the benefits of the new "appreciation" movement in music teaching (*i.e.*, the movement which seeks to add interest to musical studies through details of the lives of composers, etc.) went on to say, "What I do not see being taught or learned is the interpretative side of music as confined within the borders of the music itself."

2. Again, a leading musical critic, complaining in the *Musical Times*[2] of the low standard of certain pianoforte recitals in London wrote, "There is enough and to spare of technical ability but there seems to be a lack of intuition, cultivative imagination and real personal

[1] See the *Music Student,* Vol. VIII, 1916, p. 213.

[2] Jan., 1926.

conviction . . . Surely there must be too much time devoted to overcoming technical difficulties and far too little to cultivating the intelligent and imaginative attributes."

The evil effects of the music teacher's too great stress on technical training, and incidentally the immediate need of an adequate scheme of specific tests of musical ability for use with children, are emphasized in the following extract from a writer in the *Organist and Choirmaster*.[1] "Properly speaking, an absolutely unmusical child or person should never be allowed to learn music. Any human being who has 'no ear for music,'—*i.e.*, one who experiences no real pleasure in listening to it, and to whom there is no æsthetic difference between a finger-exercise played on the piano and the same finger-actions strummed upon the top of a table—should on no account be subjected to a course of music lessons. No sensible person would dream of boring to distraction a fellow-creature suffering from colorblindness by vainly attempting to teach him to paint—or even how to appreciate—a picture: it ought to be equally apparent that to force music-lessons upon a child suffering from tonal deafness, is only a refined form of cruelty.

"Anyone who has had much to do with local examinations in practical music knows only too well that such unmusical children do really exist. It is impossible to deny, however, that they *can* be taught in a way the externals of music: *i.e.*, they can be *made* to perform the complicated wrist and finger movements necessary for the correct and mechanical rendering of scales, arpeggios, studies and even pieces: But this undesirable consummation can only be accomplished after long months and years of hard labor and infinite patience on the part of the teacher, and of much strong crying and tears resulting in a heart-felt and life-long hatred of music on the part of the pupil. To an examiner, or to a musician-listener, the playing of such an unmusical pupil is more pathetic than pleasing: it can only tend to excite a feeling of pity for the waste of so much time and money which could have been far more usefully expended in a thousand other ways in the wide field of general technical training."

One gratifying aspect of the emphasis on interpretation rather than on technique in music is that the numerous band of prospective auditors receive due recognition at the hands of the teacher. Parker[2] rightly points out that " . . . a scheme of education which leaves

[1] April, 1915, p. 500.

[2] *Organist and Choirmaster*, Nov., 1918, p. 166.

him (*i.e.*, the auditor) out of account is not comprehensive enough . . . At bottom our musical culture very largely depends upon the intelligence and perspicuity of the average hearer. While it is a most excellent thing to have a number of highly gifted and sensitive musical people in a community, it remains true that many ambitious programs are impossible if a wide public does not prove responsive."

The conclusion of Parker's article is an effective comment on the old-fashioned systems of technical training and indicates a new field of activity for the school music teacher in a due regard for the cradling of the future musical hearers: "There are those who think that, because they can neither play nor sing, it is not worth while to trouble about music. This is a mistaken notion. It often happens that, by reason of age, occupation, or financial considerations lessons have been or are out of the question. There may, nevertheless, be a latent power of appreciation. It must not be permitted to remain undeveloped. Significant musical activity is to be found only where there exists sympathetic and enlightened appreciation, and it must not be forgotten that it is better to be a first-rate auditor than a fifth-rate executant."

It is clear therefore that a musical person may be keenly desirous of success in practical music but if the physical conditions are not favorable, technical perfection is an impossibility. Such a person may yet achieve distinction in music, say, as conductor or composer and may be able to enjoy to the full performances of what has lately been called "high-brow" music.

To sum up what has been said above: vocal and instrumental technique is so specialized a subject and the physical and physiological aspects of it are so predominant that it is not profitable for the school music teacher to concern himself with it except indirectly. The case, however, is different with interpretation.

Interpretation is the very life of *musical* as distinct from *mechanical* performance. It concerns all who are interested in any way in the art of music whether as performer, conductor, composer or intelligent listener. Moreover, interpretation is *mental* in character. As a sure basis therefore for the formation of tests of musical ability its value is evident.

COMPUTATION OF CORRELATION COEFFICIENTS

EDWARD E. CURETON

Territorial Normal and Training School, Honolulu, Hawaii

Huffaker[1] in 1925 derived a correlation formula which involved the computation of the sum of the squares of the differences between gross scores.

Orleans[2] in 1927 derived essentially the same formula, but in slightly altered form, and suggested a method for its computation. His method included the computation of the two means and standard deviations by a short method involving grouping, and the computation of either the sum of the squares of the differences between the original scores or of the standard deviation of these differences, using a new grouping interval. In either case, the grouping was not consistent throughout, and he noted that the formula was "sensitive," and subject to extremely large grouping errrors under certain conditions.

The present paper presents another derivation of this formula, a derivation of a similar formula involving the sum of the squares of the sums of gross scores, a derivation of certain checking formulæ, and a method of computation involving a system of grouping that is consistent throughout, and hence not productive of excessive grouping errors.[3] This method is especially adapted to the use of calculating machines (such as the Monroe and Merchant), and all numerical operations are positively checked.

The necessary formulæ may be derived as follows:

Let X and Y be gross scores.

x and y, the corresponding scores expressed as deviations from the respective means.

M_x and M_y, the means.

σ_x and σ_y, the standard deviations.

N, the total frequency or population.

r, the product-moment correlation coefficient.

Σ, the process of summation.

Then $r = \dfrac{\Sigma xy}{N\sigma_x\sigma_y}$ (Pearson's formula), and $\dfrac{\Sigma xy}{N} = r\sigma_x\sigma_y$

[1] A Note on Statistical Method. *Journal of Educational Psychology*, Apr., 1925, pp. 265–266.

[2] Correlation without Plotting. *Journal of Educational Psychology*, May, 1927, pp. 310–317.

[3] The writer is indebted to Dr. Hoffaker for suggesting this grouping procedure

Consider the standard deviation of a difference

$$\sigma^2_{(x-y)} = \frac{\Sigma(x-y)^2}{N} = \frac{\Sigma x^2}{N} + \frac{\Sigma y^2}{N} - \frac{2\Sigma xy}{N} = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$$

Solving for r, we have

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma^2(x-y)}{2\sigma_x\sigma_y} \tag{1}$$

Now

$$\sigma^2_{(x-y)} = \frac{\Sigma(x-y)^2}{N} = \frac{\Sigma[(X-M_x)-(Y-M_y)]^2}{N} = \frac{\Sigma[(X-Y)-(M_x-M_y)]^2}{N}$$

$$= \frac{\Sigma(X-Y)^2}{N} - \frac{2\Sigma(X-Y)(M_x-Y_y)}{N} + \frac{\Sigma(M_x-M_y)^2}{N}$$

$$= \frac{\Sigma(X-Y)^2}{N} - \frac{2\Sigma XM_x}{N} + \frac{2\Sigma XM_y}{N} + \frac{2\Sigma YM_x}{N} - \frac{2\Sigma YM_y}{N} + \frac{\Sigma M_x^2}{N} - \frac{2\Sigma M_xM_y}{N} + \frac{\Sigma M_y^2}{N}$$

Noting that $\frac{\Sigma X}{N} = M_x$, $\frac{\Sigma Y}{N} = M_y$, $\Sigma M_x = NM_x$, $\Sigma M_y = NM_y$, and $\Sigma M_xM_y = NM_xM_y$,
we have

$$\sigma^2_{(x-y)} = \frac{\Sigma(X-Y)^2}{N} - 2M_x^2 + 2M_xM_y + 2M_yM_x - 2M_y^2 + M_x^2 - 2M_xM_y + M_y^2$$

$$= \frac{\Sigma(X-Y)^2}{N} - M_x^2 + 2M_xM_y - M_y^2$$

$$= \frac{\Sigma(X-Y)^2}{N} - (M_x-M_y)^2$$

Substituting in (1),

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \frac{\Sigma(X-Y)^2}{N} + (M_x-M_y)^2}{2\sigma_x\sigma_y} \tag{2}$$

Now consider the standard deviation of a sum. By a similar line of reasoning, we have

$$r = \frac{\sigma^2(x+y) - \sigma_x^2 - \sigma_y^2}{2\sigma_x\sigma_y} = \frac{\frac{\Sigma(X+Y)^2}{N} - (M_x+M_y)^2 - \sigma_x^2 - \sigma_y^2}{2\sigma_x\sigma_y} \tag{3 and 4}$$

If we use gross-score methods of computation, the basic values needed for substitution in Formulas (2) and (4) are N, ΣX, ΣX^2, ΣY, ΣY^2, $\Sigma(X - Y)^2$ and $\Sigma(x + y)^2$. It is useful to note certain relations among these functions which may be used in checking the computations.

$$\Sigma X + \Sigma Y = \Sigma(X + Y) \tag{5}$$

Note that $\Sigma(X + Y)$ is not actually required in either Formula (2) or Formula (4), but in the method of computation outlined below it appears as a by-product in the computation of $\Sigma(X + Y)^2$, involving only a negligible amount of additional work. Note also that $\Sigma(X - Y)$ is not required. Hence in the computation of $\Sigma(X - Y)^2$ it is not necessary to take account of signs. There are thus no negatives involved in any of the preliminary computations.

From formulas (2) and (4)

$$\sigma_x^2 + \sigma_y^2 - \frac{\Sigma(X - Y)^2}{N} + (M_x - M_y)^2 = \frac{\Sigma(X + Y)^2}{N} - (M_x + M_y)^2 - \sigma_x^2 - \sigma_y^2$$

Transposing and collecting.

$$\frac{\Sigma(X - Y)^2}{N} + \frac{\Sigma(X + Y)^2}{N} = 2\sigma_x^2 + 2\sigma_y^2 + (M_x - M_y)^2 + (M_x + M_y)^2 \tag{6}$$

Also

$$\Sigma(X + Y)^2 + \Sigma(X - Y)^2 = \Sigma X^2 + 2\Sigma XY + \Sigma Y^2 + \Sigma X^2 - 2\Sigma XY + \Sigma Y^2 = 2\Sigma X^2 + 2\Sigma Y^2 \tag{7}$$

The method of computation outlined below consists essentially in substituting equivalent grouped scores for the original scores, treating these equivalent scores as gross scores, and using gross-score formula for all succeeding operations. Since we have two kinds of gross scores, the original scores and the equivalent scores. it is convenient to introduce certain modifications of the standard system of notation.

Let X and Y (or $X_0, X_1, X_2, \ldots$) be original scores.

x and y (or $x_0, x_1, x_2, \ldots$) the corresponding equivalent scores (since no deviation formulæ are used, this need not occasion any confusion).

M_x and M_y (or $M_0, M_1, M_2, \ldots$) the means of the original series.

m_x and m_y (or $m_0, m_1, m_2, \ldots$) the means of the equivalent series.

σ_x and σ_y (or $\sigma_0, \sigma_1, \sigma_2, \ldots$) the standard deviations of the original series.

s_x and s_y (or s_0, s_1, s_2, . . .) the standard deviations of the equivalent series.

AO, the arbitrary origin, or the score in an original series corresponding to a score of zero in the equivalent series.

i, the width of a grouping interval.

d, a general term for x or y (or x_0, or x_1, or x_2, . . .) used in computing the mean and standard deviation of a particular equivalent series. (d is a deviation from the AO in terms of the equivalent-score units, but since the AO will always be taken at the zero of the equivalent series, this then becomes the same as an equivalent score. The use here of the term "deviation" does not imply the use of a deviation formula, where deviations are taken from the mean of the series.)

N the total frequency or population.

f, the frequency of a score-group.

Σ, the process of summation.

v, the second moment about an AO in terms of the equivalent-score units $\left(v = \frac{\Sigma d^2}{n}\right)$.

n, the numerator of the correlation coefficient when computed by Formula (2) or (4) $\left(\text{so that } r = \frac{n}{2s_x\, s_y}\right)$

c, the correction for grouping to the mean of an original series ($c = im$).

r, the correlation coefficient.

The steps in computation are as follows:

1. Arrange the original scores in columns, the order of the subjects in each column being the same (usually alphabetical), on a sheet of paper. We will call this the original data sheet. This sheet should ordinarily contain in its heading the title of the study, the subjects, the investigator, the computer, the date, and a description of each of the variables. Some investigators include also the means, standard deviations and correlation coefficients, blanks being left for this purpose.

If there are only two variables, these may be labeled X and Y. If there are a number of variables, they should be numbered. Two variables may, of course, be numbered; this has been done here in the accompanying illustration in order to show the general method.

A blank space should be left at the right of each column of original scores for the recording of the corresponding equivalent scores.

An alternative method to the above is to record all the data for each subject on a card.

2. In each column note the highest score and the lowest; find the difference; divide this difference by twelve, dropping any fractional remainder; and select this or the next lowest convenient number as the grouping interval (i). (See Table I.) Convenient intervals are 1, 2, 5, 10, 20, 50, etc. If this method is adhered to, there will be from twelve to thirty categories, and the grouping errors will not ordinarily affect the standard deviation by more than one per cent.

3. Prepare a standard deviation sheet for each variable. This sheet need be headed only with the number of the variable and the title of the study or key-number of the original data sheet.

Record in the two left-hand columns the limits of the grouping intervals (See Tables II and III.) The interval at the top should contain the highest score, and each interval should be of the width determined on the original data sheet (i). The intervals are recorded on successive lines until one is reached which contains the lowest score.

Blank columns are left for the tabulation and frequency records. In the two right-hand columns (headed d and d^2), are recorded the equivalent scores and their squares. Column d should always begin with zero at the bottom and progress upward in single steps.

4. Tabulate the scores for each variable on the appropriate standard deviation sheet, and record the equivalent scores on the original data sheet. To do this, it is necessary to read each original score; tabulate it in the proper group on the standard deviation sheet; and record the corresponding entry in the d column of the standard deviation sheet, at the right of the original score on the original data sheet. By this process, the tabulation for the mean and standard deviation, and the assigning of equivalent scores are done for each variable at one operation.

This operation should be performed with great care, and if absolute accuracy is desired, it should be done twice, reading the scores on the original data sheet once from the top down and once from the bottom up. In the second reading it is not necessary to record the tabulation, but only to check the equivalent scores. If the latter are correct, any errors in tabulation (which are not very likely in this case) will be discovered in subsequent checks. If an error is found, the tabulation must of course be corrected along with the equivalent score.

TABLE I.—ORIGINAL DATA SHEET

Name	Variable 1		Variable 2	
	X_1	x_1	X_2	x_2
A	83	4	64	2
B	125	8	93	5
C	72	3	112	7
D	84	4	72	3
E	96	5	93	5
F	83	4	45	0
G	192	15	184	14
H	49	0	58	1
I	108	6	126	8
J	132	9	98	5
K	148	10	144	10
L	98	5	73	3
M	113	7	122	8
N	184	14	171	13
O	63	2	46	0
P	157	11	189	14
Q	161	12	127	8
R	89	4	67	2
S	172	13	173	13
T	53	1	42	0
U	183	14	177	13
V	91	5	102	6
W	113	7	87	4
X	83	4	121	8
Highest	192		189	
Lowest	49		42	
Difference	143		147	
Dif./12	11+		12+	
i	10		10	

5. Record on each standard deviation sheet the values AO and i. The AO is the mid-point of the lowest interval. Thus for Table II, AO $= \frac{40 + 49}{2} = 44.5$; and for Table III, we happen in this case to have the same value, though this is not usually the case. The value of i is taken from the original data sheet, or by inspection, from the two left-hand columns of the standard deviation sheet.

Record in the f-column the summary of the tabulation.

TABLE II.—STANDARD DEVIATION SHEET
Variable 1

From	To	Tabulation	f	d	d^2
190	199	1	1	15	225
180	189	11	2	14	196
170	179	1	1	13	169
160	169	1	1	12	144
150	159	1	1	11	121
140	149	1	1	10	100
130	139	1	1	9	81
120	129	1	1	8	64
110	119	11	2	7	49
100	109	1	1	6	36
90	99	111	3	5	25
80	89	IIII	5	4	16
70	79	1	1	3	9
60	69	1	1	2	4
50	59	1	1	1	1
40	49	1	1	0	0

AO = 44.5.
i = 10.
N = 24.

Σd = 167
Σd^2 = 1599.

Data sheet (Table I)

Computations

Operation	Value	Result
$\frac{\Sigma d}{N}$	6.958	m_1
$\frac{\Sigma d^2}{N}$	66.625	v
$m \times m$	48.414	m^2
$v - m^2$	18.211	s^2_1
$\sqrt{s^2}$	4.27	s_1
$i \times m$	69.58	c
$c + \text{AO}$	114.08	M_1
$i \times s$	42.7	σ_1

(Do the last three computations twice.)

Compute N, Σd, and Σd^2. This is performed on the calculating machine as follows: Divide the key-board into two sections. The section at the left should contain two columns, and the section at the right the rest of the keyboard. Punch the first value in the d-column in the left-hand section, the corresponding value in the d^2-column in the right-hand section, and setting the machine for multiplication, turn in these numbers to the value of the corresponding entry in the f-column. By this process we multiply d and d^2 by f simultaneously. Clear the key-board, *but not the dials.* Punch the second value in the d-column in the left-hand section, the second value in the d^2-column in the right-hand section, and multiply by the second value in the f-column. Clear the key-board and proceed with the third set of values, continuing to the bottom of the column (or the top; the direction is immaterial); not omitting to turn in with cleared key-board the lowest value in the f-column. Carriage shifting and all the usual short-cuts may be employed in multiplying by each value of f. At the conclusion of the operation, Σd will appear in the left-hand section of the lower dial, Σd^2 in the right-hand section, and N in the upper dial.

If the values of f are large, it may be wise to clear the upper dial at each step, checking each f with the number appearing in the upper dial. In this case, N will have to be computed separately.

If a nine or ten-bank machine is available, it may be divided into three sections, with one column at the left, three or four in the center, and five at the right. At each step now, we punch in the value one in the left-hand column (this may be done once for all by setting the item-count if the machine has one). d in the center column, and d^2 in the right-hand column; multiplying by f as before. The upper dial may be cleared at each step, and at the conclusion of the operation the lower dial will show N at the left, Σd in the center, and Σd^2 at the right. This method is only applicable, of course, where the population is not extremely large, so that Σd^2 does not contain more than five figures, and Σd does not contain more than three or four.

Record the values of N, Σd, and Σd^2 on the standard deviation sheet. (See Tables II and III.)

6. Prepare for each pair of variables which it is desired to correlate, a supplementary correlation sheet. This should be headed with the number of the x-variable, the number of the y-variable, and the title of the study or key-number of the original data sheet. Columns should be provided for tabulating the $(x + y)$ and $(x - y)$ values.

TABLE III.—STANDARD DEVIATION SHEET
Variable 2

From	To	Tabulation	f	d	d^2
180	189	11	2	14	196
170	179	111	3	13	169
160	169	....	.	12	144
150	159	....	.	11	121
140	149	1	1	10	100
130	139	....	.	9	81
120	129	1111	4	8	64
110	119	1	1	7	49
100	109	1	1	6	36
90	99	111	3	5	25
80	89	1	1	4	16
70	79	11	2	3	9
60	69	11	2	2	4
50	59	1	1	1	1
40	49	111	3	0	0

AO = 44.5.
i = 10.
N = 24.

Σd = 152
Σd^2 = 1458.

Data sheet (Table I)

Computations

Operation	Value	Result	
$\frac{\Sigma d}{N}$	6.333	m_2	
$\frac{\Sigma d^2}{N}$	60.75	v	
$m \times m$	40.107	m^2	
$v - m^2$	20.643	s_2^2	
$\sqrt{s^2}$	4.54	s_2	
$i \times m$	63.33	c	
$c + \text{AO}$	107.83	M_2	
$i \times s$	45.4	σ_2	

(Do the last three computations twice.)

The former should extend from zero to the sum of the maximum d-values on the standard deviation sheets. The latter should extend from zero to the higher of the two maximum d-values. (See Table IV.)

Tabulate the $(x + y)$ and $(x - y)$ values. This is done as follows: Going back to the original data sheet, take the sum and difference of the equivalent scores, in the two variables under consideration, for each subject. The sum is entered in the $(x + y)$ tabulation; the difference in the $(x - y)$ tabulation. In recording the difference no account is taken as to sign, the smaller value being subtracted from the larger in each case.

When this tabulation is finished, compute $\Sigma(x + y)$ and $\Sigma(x + y)^2$ in the same manner as Σd and Σd^2 were computed on the standard deviation sheets. It is not necessary to compute N again, though this may be done as a partial check. Compute $\Sigma(x - y)^2$ in a similar manner. We cannot, of course, compute $\Sigma(x - y)$, as signs were disregarded in tabulating, and there is no reason for computing $\Sigma|(x - y)|$.

Record $\Sigma(x + y)$, $\Sigma(x + y)^2$, and $\Sigma(x - y)^2$ on the supplementary correlation sheet. Copy from the Σd and Σd^2 values on the standard deviation sheets, Σx, Σy, Σx^2, and Σy^2. (See Table IV.)

Now make the preliminary checks. These are illustrated in Table IV at the top of the three right-hand columns. The first-degree terms are checked by the formula $\Sigma x + \Sigma y = \Sigma(x + y)$. The second-degree terms are checked by the formula $2\Sigma x^2 + 2\Sigma y^2 = \Sigma(x + y)^2 + \Sigma(x - y)^2$. (Formulas (5) and (7) in the derivation.) If the values of Σx, Σy, Σx^2, Σy^2, $\Sigma(x + y)$, $\Sigma(x + y)^2$, and $\Sigma(x - y)^2$, when substituted in these equations, are found to lead to identities, the preliminary computations may be considered to have been correctly performed. If they do not, these values must be re-computed until the error is discovered and rectified.[1]

7. Compute on each standard deviation sheet the values m, s^2, and s. These computations are illustrated in the three right-hand columns in Tables II and III. These values should be labelled with subscripts denoting the number of the variable.

8. Copy on the supplementary correlation sheet the values of m_x, m_y, s_x^2, s_y^2, s_x, s_y, and N from the standard deviation sheets. (See Table IV.) Compute the value of r. This computation is illustrated in Table IV in the three right-hand columns. Note that if an error

[1] This method is adapted from the one devised by Kelley for checking the computations on his Correlation Chart. See Kelley, T. L.: "The Interpretation of Educational Measurements." World Book Co., 1927, pp. 163–169.

TABLE IV.—SUPPLEMENTARY CORRELATION SHEET
x-variable 1, y-variable 2, Data sheet (Table I)

$(x+y)$	Tabulation	f	$(x+y)^2$
29	1	1	841
28	..	.	784
27	11	2	729
26	1	1	676
25	1	1	625
24	..	.	576
23	..	.	529
22	..	.	484
21	..	.	441
20	11	2	400
19	..	.	361
18	..	.	324
17	..	.	289
16	..	.	256
15	1	1	225
14	11	2	196
13	1	1	169
12	1	1	144
11	11	2	121
10	11	2	100
9	..	.	81
8	1	1	64
7	1	1	49
6	11	2	36
5	..	.	25
4	1	1	16
3	..	.	9
2	1	1	4
1	11	2	1
0	..	.	0

$E(x+y) = 319.$
$E(x+y)^2 = 5979.$

$(x-y)$	Tabulation	f	$(x-y)^2$
15	.....	.	225
14	.....	.	196
13	.....	.	169
12	.....	.	144
11	.....	.	121
10	.....	.	100
9	.....	.	81
8	.....	.	64
7	.....	.	49
6	.....	.	36
5	.....	.	25
4	𝍸	5	16
3	111	3	9
2	𝍸	5	4
1	𝍸 111	8	1
0	111	3	0

$\Sigma(x-y)^2 = 135.$

$\Sigma x = 167.$
$\Sigma y = 152.$
$\Sigma x^2 = 1599.$
$\Sigma y^2 = 1458.$

$m_x = 6.958.$
$m_y = 6.333.$
$s_x^2 = 18.211.$
$s_x = 4.27.$
$s_y^2 = 20.643.$
$s_y = 4.54.$
$N = 24.$

Preliminary checks		
Operation	Value	Check
$\Sigma x + \Sigma y$	319	$\Sigma(x+y)$
$2\Sigma x^2 + 2\Sigma y^2$	6114	$\Sigma(x+y)^2 + \Sigma(x-y)^2$
Computations		
$\frac{\Sigma(x+y)^2}{N}$	249.125	$v_{(x+y)}$
$\frac{\Sigma(x-y)^2}{N}$	5.625	$v_{(x-y)}$
$m_x + m_y$	13.291	$(m_x + m_y)$
$(m_x + m_y) \times (m_x + m_y)$	176.651	$(m_x + m_y)^2$
$m_x - m_y$	.625	$(m_x - m_y)$
$(m_x - m_y) \times (m_x - m_y)$	.391	$(m_x - m_y)^2$
$2s_x^2 + 2s_y^2 + (m_x + m_y)^2 + (m_x - m_y)^2$	..... 254.570	$v_{(x+y)} + v_{(x-y)}$ (check)
$\sqrt{4s_x^2 s_y^2}$	38.77	$2s_x s_y$ (check)
$v_{(x+y)} - (m_x + m_y)^2 - s_x^2 - s_y^2$	33.620	n
$s_x^2 + s_y^2 + (m_x - m_y)^2 - v_{(x-y)}$	33.620	n (check)
$\frac{n}{2s_x s_y}$	.87	r

(Do the last computation twice.)

TABLE V.—SUPPLEMENTARY CORRELATION SHEET, SHORTER FORM
x-variable 1, y-variable 2, Data sheet (Table I)

$(x-y)$	Tabulation	f	$(x-y)^2$
15	.....	.	225
14	.....	.	196
13	.....	.	169
12	.....	.	144
11	.....	.	121
10	.....	.	100
9	.....	.	81
8	.....	.	64
7	.....	.	49
6	.....	.	36
5	.....	.	25
4	𝍸	5	16
3	111	3	9
2	𝍸	5	4
1	𝍸 111	8	1
0	111	3	0

$\Sigma(x-y)^2 = 135.$

$N = 24.$
$m_x = 6.958.$
$m_y = 6.333.$
$s_x^2 = 18.211.$
$s_x = 4.27.$
$s_y^2 = 20.643.$
$s_y = 4.54.$

Computations		
Operation	Value	Result
$\frac{\Sigma(x-y)^2}{N}$	5.625	$v_{(x-y)}$
$m_x - m_y$	.625	$(m_x - m_y)$
$(m_x - m_y) \times (m_x - m_y)$	.391	$(m_x - m_y)^2$
$s_x^2 + s_y^2 + (m_x - m_y)^2 - v_{(x-y)}$	33.620	n
$2 \times s_x \times s_y$	38.77	$2s_xs_y$
$\frac{n}{2s_xs_y}$	.87	r

is indicated at the first check (Formula (6) in the derivation), this error may be on the supplementary correlation sheet or on either of the standard deviation sheets.

9. Compute on the standard deviation sheets the values of M and σ. These computations are illustrated in Tables II and III in the three right-hand columns, at the bottom. This completes the work.

If the above procedure is followed exactly, the result will give the means, standard deviations, and correlations for any number of variables, and the probability of numerical errors in the final results will be extremely small.

If speed of computation is more important in any particular case than extreme accuracy, the method may be materially shortened by sacrificing the system of numerical checks. Going back to the beginning, steps one, two, and three are performed as indicated. In step four, the assigning of equivalent scores is done only once. Step five is performed as indicated, followed at once by step seven and then step nine. A special supplementary correlation sheet is now prepared for each pair of variables which it is desired to correlate. (See Table V.) On this sheet only the $(x - y)$ values are tabulated. Now compute $\Sigma(x - y)^2$; copy from the standard deviation sheets the values of N, m_x, m_y, s_x^2, s_y^2, s_x, and s_y; and compute r as illustrated in the three computation columns of Table V. With the use of a calculating machine, this is the most rapid method for computing means, standard deviations and correlation coefficients with which the writer is acquainted.

If the population is small, a still further saving of the time may sometimes be effected by computing $\Sigma(x - y)^2$ directly. To do this, set the machine to add, and subtracting the corresponding equivalent scores on the original data sheet, square each difference mentally and put this value in the machine. At the end of this operation, $\Sigma(x - y)^2$ will appear immediately as the sum. To use this method, the computer must know the squares, but if the correlation is positive, and particularly if it is high, the differences will not in general be large; and in any case there will never be a difference greater than thirty.

The method of computing means and standard deviations outlined herein may be applied to single-variable problems as well as to those involving correlation.

A friend of the writer's worked a correlation by four different methods, with the following results:

1. Kelley Correlation Chart, one hour, twenty-five minutes. Gives complete check and scattergram, permitting visual inspection for non-linearity.

2. Writer's long method, one hour, five minutes. Gives complete check but no scattergram.

3. Writer's short method, fifty minutes. Gives neither check nor scattergram.

4. Ordinary scattergram method, taking arbitrary origins at zero, fifty-five minutes. Gives partial check, but does not give means nor standard deviations without additional work.

A ten-bank Monroe Calculator was used for all calculations. There were one hundred thirty-seven cases, and the correlation was . 83 in the problem.[1]

Summary

A method has been devised which by the omission of the scattergram and the use of a calculating machine, materially shortens the time required to compute and check a correlation coefficient and its attendent means and standard deviations. Every step is mechanical so that a clerk who knows nothing of statistics can be taught to use the method. The resulting coefficient is a "true" product-moment coefficient, having exactly the same numerical value as one computed by the standard Pearson formula, using the same grouping. The method effects a still greater saving of time when there are a number of inter-correlations to be computed, as none of the labor of computing means and standard deviations has to be repeated.

[1] The writer is indebted to Mr. P. J. Rulon of Stanford University for making this test.

EXAMINATION AS AN AID TO LEARNING

ARTHUR T. JERSILD

Barnard College, Columbia University

The value of classroom examination as an aid to learning is twofold.[1] One effect of examination is to compel the student to rehearse and recapitulate what he has already learned. The advantage of exercise of this kind speaks for itself.[2] Another advantage arises from the influence which examination has on the learner's attitude. When put to a practical test, the learner is called upon to work under pressure, to participate more actively in the learning situation. It is with this latter feature that the present investigation is concerned.

The question here raised is this: What benefit derives from the examination method apart from the advantages gained through the exercise of associations that have been formed during previous periods of study? To what extent does examination tend to induce an attitude that is favorable to more effective learning and recall? The question so raised presents an immediate difficulty, for it is obviously impossible to devise an examination which involves much attitude and no exercise. But it is possible to examine the student on materials that are entirely new to him and on which he has formed a minimum of previous associations, and then to observe whether the mental activity enforced during the examination period yields higher returns than the exercise which takes place under ordinary conditions of study.

In this investigation the equivalent-group method of experimentation was used. In each experiment two groups of students were given the same learning assignment. One group was given a preliminary test (pre-examination) covering the assignment before any time had been given for study and before any previous information had been divulged. The other group, which served as control, was not pre-examined. A description of the various experiments follows.

---

[1] An extensive study of the merits of classroom examination has been made by H. E. Jones: "Experimental Studies of College Teaching." Archives of Psychology, No. 68, 1923. Jones used the method of subjecting his students to brief examinations at the end of class lectures. On later tests, conducted over intervals of from three days to eight weeks, the effect was to produce scores more than fifty per cent higher than when no examination had been given.

[2] This point has been examined carefully by A. I. Gates, among others. "Recitation as a Factor in Memorizing." Archives of Psychology, No. 40, 1917.

I. The experimenter had at his constant disposal two sections of a class in beginning psychology. At the first meeting of one of these sections, on the first day of the school year, the hour was devoted to a true-false test. This examination covered materials that would be treated in lectures and reading assignments during the first three weeks of school.[1] The second section of the class was not thus pre-examined, but was given a regular lecture. On the day set for the first regular examination, the two sections were given the true-false test which had previously been administered to only one of the groups. In other words, the pre-examined group responded to a test which it had been given before, while the control group now met this test for the first time.

II. A true-false test was again used for pre-examination in connection with a collateral reading assignment. On the day when this assignment (Hart's "Psychology of Insanity") was made, and before any students had formed an acquaintanceship with the book, a true-false test was given to one section of the class, in this instance the section which had not been pre-examined in the first experiment. On the date scheduled for examination on the outside reading, this same test was given to both sections of the class.

III. In subsequent experiments, the learning exercise consisted of selections prepared for oral class presentation. The first of these was an essay on the subject of reaction-time. This topic was chosen because of its definite character, which lends itself well to scoring on an objective basis in recall tests, and for the reason that students in beginning psychology have practically no knowledge of the detailed information which is available on the subject.

A special form of pre-examination was devised for this experiment. The points contained in the essay were reduced to summary statements and framed into multiple-choice questions. No misinformation was conveyed in these questions, but each one did contain an alternative which could be correctly chosen as the truest answer to the question involved. Here is an example:

Was the reaction-time experiment primarily designed to measure (1) the speed of nerve conduction; or (2) the speed of the processes, both nervous and mental, which are the conditions of conscious life?

[1] On this first test the average score of the group was a minus quantity, indicating that the test not only touched upon unfamiliar topics but revealed a good deal of misinformation besides. The following proposition, for example, elicited more than a chance number of errors: "The speed of nerve conduction roughly corresponds to the speed of light." This was marked *true* by three-fourths of the class.

According to the historical definition, alternative (2) is the correct answer. Another example:

Which of the following four hypotheses best accounts for the finding that auditory reaction-time is quicker than visual reaction-time? (Then followed a summary statement of each of four theories that have been proposed.)

Ten questions, covering the entire exercise, were presented in this way. In each case the student was to indicate in writing the number of the alternative which he chose.

It can readily be seen that questions of this kind convey a good deal of advance information. To compensate this, the control group was also given a preliminary exercise. This consisted of a summary of all the points contained in the essay, presenting, in effect, an answer to each of the questions raised in the pre-examination given to the experimental group.

Following the pre-examination and the presentation of the preliminary summary the selection was read to each of the respective groups. Immediately thereafter two final tests were given. One of these was a test for recall, presented on prepared forms, and calling for detailed answers. Another was a true-false test. The questions used for pre-examination were not repeated in these final tests.

After the elapse of four weeks both groups were given a second test for recall. This was substantially a repetition of the recall test which had been given immediately after the reading of the reaction-time essay.

IV. The methods employed in the foregoing experiment were used a second time in an experiment conducted with two sections of another class in beginning psychology. The learning exercise in this case again consisted of an essay on the subject of reaction-time. The multiple-choice method of pre-examination was repeated, as was the presentation of a preliminary summary. In this instance, however, the multiple-choice questions comprising the pre-examination, and the preliminary summary were presented to the students in writing on multigraphed sheets. The control group was directed to study the summary during the time allowed, and the pre-examined group was told to answer the questions. Both sections were again given the same amount of time for this preliminary performance. Then followed the oral presentation of the reaction-time essay, and immediately thereafter each group was given a test for recall, again presented on prepared forms and calling for detailed answers, and a final true-

false test. In this experiment, as in the foregoing, the questions used for pre-examination were not used in the final tests.

V. A variation in technique was introduced in connection with another exercise prepared for oral class presentation. The learning selection in this case consisted of seventy narrative statements, all bearing upon the biography of a fictitious character. The two sections of the class in beginning psychology that served in previous experiments, and four sections of a class in experimental psychology served as subjects.

The pre-examination questions used in this experiment called for written essay answers. The subject-matter of the biography was broken up into topics, and on the basis of these topics questions were asked. The students could answer the questions only by guessing, since no specific information had been given prior to the pre-examination. Following are sample questions:

What was the character of the man's early home environment?
What characteristics distinguished him as a child?
What were his main interests at high school?
At college?
What vocation did he enter? etc.

In answering these questions the subjects were compelled to write at a great speed, since the questions were presented in rapid succession.

Even though no specific information is conveyed by a pre-examination of this type, yet it can be seen that the questions reveal the general framework of the biography. This fact was taken into account by giving the control group the same amount of general information in the form of a preliminary topical outline: the information was given that a biography would be read; that this would be followed by a recall test; that the biography touched upon the man's early home surroundings, his characteristics as a child, his interests, his activities at school, etc. The control group was required merely to listen to this preliminary introduction, and no questions were asked. The preliminary period for both groups was of the same amount of time.

Following this, the biography was read. Immediately after the completion of the reading the students were told to write from memory as many statements from the selection as they had retained. The score was the number of statements correctly recalled.

This completes the description of procedure and method. It will be observed that in each of the foregoing experiments the effect of the

experimental factor must be found in the difference between the average score made by the pre-examined group and that made by the control group. Two precautions were taken to rule out, as far as possible, any constant error that might arise from the use of the equivalent group method. First of all, the members of each group were paired with members of the other on the basis of scores in the general college entrance intelligence test. Exact point-for-point pairing was in some cases impossible, but the differences were very slight and were balanced between the groups. Secondly, each of the sections of the class in beginning psychology that served in all the experiments save the third, functioned alternately as pre-examined group and as control group in the successive experiments.

We come now to the results. For the sake of brevity and clarity all of the results have been reduced to a common denominator and have been brought together into one table. This table gives for each experiment the score made by the pre-examined group, that made by the control group, and, in parenthesis, the reliability of the difference between these scores. In deriving the reliability of the difference it was necessary to obtain: The average score made by each group in the tests that followed the learning period; the SD of the distributions represented by these averages; the SD of each separate average; the difference between the two averages; the SD of this difference; the quotient obtained by dividing the difference by the SD of the difference; and, from this, the reliability of the difference in terms of the number of chances in 100 that the actual difference will be greater than zero.[1] The method used in reducing all the averages to a common denominator was: Divide the pre-examined group average and the control group average by the magnitude of the control group average, and multiply the two quotients so obtained by 100 (*e.g.*, the former average is 15, latter twelve; the ratio is 15:12; divided by 12, giving the ratio 1.25:1; multiply by 100, giving 125:100). The reliability of the difference between the averages was, of course, derived from the actual averages and not from these ratios.

Key to Table

I. Classroom lectures and reading assignments. True-false preexamination. Same true-false test used for final examination. Thirty-seven subjects in each group.

[1] The formulæ used in deriving these measures were taken from Garrett: "Statistics in Psychology and Education." New York: Longmans, Green and Co., 1926, p. 135, and pp. 145–146.

II. Collateral reading assignment. True-false pre-examination; same test used for final examination. Forty-two in each group.

III. Reaction-time essay, read orally. Multiple-choice questions on pre-examination; control group given a preliminary oral summary. Final tests (1) recall; (2) true-false; (3) recall test after elapse of four weeks. Forty-two in each group.

IV. Second reaction-time exercise; pre-examination and preliminary summary presented in writing. Final tests (1) recall; (2) true-false. Thirty-five in each group.

V. Biography selection. Essay pre-examination; control group given a preliminary topical outline. Final test, straight recall. Sixty-three in each group.

	I	II	III			IV		V
			(1)	(2)	(3)	(1)	(2)	
Pre-examination.....	97.5	102.5	113	115	114	120	105	105.2
Control.....	100	100	100	100	100	100	100	100
Reliability.....	(64)	(77)	(97)	(100)	(81)	(80)	(84)	(96)

The first two entries (I and II) in the table show a somewhat surprising result. In each of these experiments the pre-examined group had been given at a previous time the true-false test on which the results in the table are based. In one instance (I) the pre-examined group made a lower score than did the control group which had never met the test before. In the second case (II) the pre-examined group is slightly in the lead, but the difference between the two averages is small, and the reliability is low. In other words, pre-examination with a true-false test demonstrates partly negative and essentially unreliable advantages for the pre-examination method.

A different result appears from the experiments in which other methods of pre-examination were used. The experiments in which multiple-choice and essay questions comprised the pre-examination show a consistently higher score for the pre-examined group (entries III, IV, V). The advantage ranges from five per cent to twenty per cent. The reliability in these instances, although not consistently high, is sufficiently pronounced to indicate a true difference in favor of the group that was pre-examined, and to demonstrate the superiority of these methods of examination.

The results thus summarized deserve further comment. They present, in the first place, an indictment of the true-false test. So little did the students profit from the exercise of responding to a test

of this kind that on a second attempt, after an interval of study, they were no better off than the students who had never met the test before. It appears that the true-false test is of dubious value as a pedagogical instrument.[1] On purely theoretical grounds it is open to two serious charges. First, a test of this kind, presenting as it does a random and unpredictable intermingling of true and false propositions, may have just as much the effect of perpetuating error as of strengthening proper associations and stimulating wholesome curiosity. Each statement calls for a categorical *true* or *false*. Whether the response is right or wrong the mere act of putting a stamp of affirmation or denial on a given statement has the effect of strengthening the association so formed. If the test is designed to be at all diagnostic of the relative standing of the members of a class it will necessarily evoke a large number of erroneous judgments on the part of the majority of the students. Positive benefit will arise from the exercise of a correct response, either in affirming a proposition that is true or in registering denial of one that is false; but this in turn may be offset by the disadvantage arising from the exercise of an incorrect response, either in subscribing to a false proposition or in stamping as false one that is really true. If the test has been used for preexamination purposes the wrong associations will be quite as strong as those that are right. The law of primacy in persuasion will here be operative at full force. When the student later undertakes the study of the subject-matter on which he was tested he does so under the influence of preconceived notions. If the points that are met in subsequent study stand out in sharp denial of his previous opinions the effect will no doubt be salutary. But if the issue is somewhat obscure and demands careful scrutiny the student can readily take recourse to his previous verdict and let well enough alone.

Another shortcoming of the true-false test is the fact that it does not make strong demands upon the industry of the examinee. On the whole, it is more nearly a test of passive recognition than of active recall. In responding to a true-false test the student is not called upon to organize his knowledge or to reduce it to systematic statement with proper emphasis on the most significant details. He need only indicate his acceptance or denial of each of the successive propositions. When once he has committed himself, the validity of his initial judg-

[1] This criticism is directed against the true-false test only insofar as the test should serve as an aid to learning. Its merits as an expeditious means of testing the student's ready information and of determining his relative standing in the class are not here called into question.

ment will seem all the more compelling. The test does not compel an attitude of intense participation in the learning task.

The same criticisms cannot be urged against examination by means of specific interrogatives in the form of multiple-choice or essay questions. A direct question neither sanctions error nor invites composure, but comes rather as a challenge to the individual's attention and as an instigator of doubt and perplexity. We have seen that the mental activity enforced during a brief period of examination by means of specific questions is conducive to more effective learning in a subsequent period of study than is the mental activity involved in attending to narrative statements contained in a preliminary summary or a true-false test. It appears from this result that a direct interrogation constitutes a more intense stimulus than does a narrative statement, and will, accordingly, give rise to a more lively response; and that an examination serves as an aid to learning insofar as it puts this principle to a practical account by stimulating the industry of the learner.

A STUDY OF THE EFFECT OF A DEFINITE COURSE OF READING IN GENERAL LITERATURE UPON ACHIEVEMENT IN CONTENT SUBJECTS WITH CHILDREN OF SUPERIOR MENTAL ABILITY

CORA LEE DANIELSON

Department of Psychology and Educational Research of the Los Angeles City Schools

The reading of children of superior mental ability has been subjected to research studies fairly numerous and somewhat varied in nature, although the major part of these studies have concerned themselves with the type and quantity of the selected reading matter rather than with the effectiveness of the reading. No investigator seems either to have confirmed or to have rejected the conclusions of the Standard University investigation as they are expressed by James De Voss in "Genetic Studies of Genius."

The particular conclusions referred to are in the following quotation:

"It may be stated that the achievement level of the gifted children is forty per cent above that expected for their chronological ages and approximately ten per cent below that expected for their intelligence level . . . While the gifted children have an achievement level well above that of other children, it is never-the-less so much below their level of intelligence as to make the condition a definite challenge to both the parent and to the teacher."[1]

Although any doubt as to the authenticity of the data or of the statistical results upon which the conclusions are based is precluded by the knowledge of the source and of the purpose of the study from which the quotation is copied, some pertinent questions are actuated by the statement. It is permissible to inquire out of what material and by what method the bridge is to be built across the gap between the high intelligence level and the low educational level. The undesirable condition may offer a challenge to the parent and to the teacher which they may accept, but they must look to the expert in education for a plan of procedure that will be effective in removing the condition. Does the expert know what the procedure is to be? Does it depend upon teaching material? Will a child of a superior level of intelligence maintain a corresponding educational level if he is exposed to reading material of an appropriate quality and quantity? Or is the difficulty

[1] Terman, L. M.: "Genetic Studies of Genius." Palo Alto: Stanford University Press, 1925, Vol. I, p. 347.

one of unsuitable teaching method? If so, what method has been proven more efficacious than those methods generally pursued?

The search for light upon these questions prompted the writer to submit certain available test data to statistical treatment in order to determine the change in achievement in content subjects following a definitely planned course in reading which is one of the phases of curriculum enrichment for children of superior mental ability in segregated groups in the elementary schools of the Los Angeles City School System.

The Course in Reading

The purpose of the course of study in reading, stated in general terms, is strengthen the habit of regular purposeful reading and to develop an appreciation of literature. More specifically the objects are:

1. To provide interesting material commensurate with the mental ability of the pupils.
2. To expose the pupils to selections of material so rich and varied that narrowed interests will not become fixed.
3. To encourage reading habits that will carry over into the home life of the pupils.

The Procedure

For our purpose reading matter is classified rather arbitrarily nto seven fields: Drama, fiction, history and biography, mythology and folklore, periodicals, science and invention, sociology and travel. Successive periods of approximately six weeks are devoted to reading in each of these fields. Deposits of books are obtained from the city school library and from the public libraries. In addition various well-selected annotated book lists are furnished. From the books or from the annotated lists of titles each child makes his own selection, reads when and where he wishes and, with the class as an audience, contributes to the program a report, a book review, the reading of a well liked portion, a dramatization, or what not. The nature of the pupil "contribution" depends upon the nature of the material and the wishes of the pupil. The essential is the doing of something with the material that has been read. A minimum amount of reading that shall be accomplished by each pupil in each period of time is in the minds of the teachers and of the supervisor. However, the necessity of announcing the minimum to the pupils does not arise.

Rather does it seem necessary to establish a maximum amount because time prevents further reporting.

At the end of five or six weeks the books on science and invention, perhaps, are replaced by books of fiction, or of history and biography, and a type of pupil expression suitable to the new material is decided upon and maintained to make use of the information gained through the reading. The wide variety of books in each of the classifications, and the great amount of informal talking about the books insures the interest of each individual. During the course of a year he is acquainted with a wide field of literature. Diaries of home reading kept by the pupils and reports made by librarians indicate that the acquaintance with a variety of good reading material is continued.

Subjective evidence such as diaries and reports may be accepted as proof of the effectiveness of the particular course of study in creating desirable habits in reading and appreciation of literature. However, more objective evidence of the possible relation between the particular course and the charge in the achievement level of the pupils is necessary if we are to determine to what extent the method of teaching meets the challenge mentioned by De Voss.

For objective evidence we are depending upon data obtained from Stanford Achievement Tests. We are considering the change of achievement in purely informational or content subjects since they are probably more directly influenced by reading than are skill subjects. That is, reading rate and reading comprehension, arithmetic, language usage and spelling are the direct product of drill. Therefore, gain in the test scores in these subjects would not be acceptable as evidence of the result of much varied reading. The three tests of the Stanford Achievement battery that would seem to measure the results of reading in terms of gain in knowledge of facts are Test 6, a general information test in nature study and science, and Test 7, a general information test in history and literature. In addition gain in "Total educational age" has been considered. Total educational age as used in this study is the educational age of the complete Stanford Achievement Test. It is the average for the nine separate tests that make up the battery. It therefore includes seven tests of skill in addition to the two tests of general information. It is derived from a composite score the reliability of which is acknowledged to be greater than of the separate subjects. Its use in this study is not essential, but a comparison between the growth of mental age and the growth of educational age is of interest.

Statistical Evidence

At the end of June, 1927 the complete test data on two hundred twenty-seven pupils were available for this study. These data are the results of intelligence tests and of Stanford Achievement Tests. The Stanford-Binet and an initial Stanford Achievement were given to each pupil at the time he was selected for segregation in the curriculum enrichment rooms. The second Achievement Test was given in June, 1927. This divided the two hundred twenty-seven pupils into four groups dependent upon the interval between the initial and the second Achievement Test. Group I consisted of fifty-eight pupils with a one-half year interval between the tests. Group II was composed of ninety-four pupils with a one year interval between the tests. Group III contained forty-five pupils with a one and a half year interval between tests, and Group IV consisted of thirty pupils with approximately two years between the tests. In each group the range of intelligence, from 125 IQ to 150 (and above), was divided into four intervals of ten points.

The gain in mental ages and in educational ages in five school months is shown for Group I in Table I.

Table I.—Gains in Mental Ages and in Educational Ages of 58 Pupils (Group I) of Superior Mental Ability Measured by Tests Given One-half Year Apart

IQ	N	MA			EA in nature study and science			EA in history and literature			Total EA		
		Test I	Test II	Gain	Test I	Test II	Gain	Test I	Test II	Gain	Test I	Test II	Gain
125–130	14	175	181	6	169	171	2	172	183	11	176	185	9
130–139	23	178	183	5	170	175	5	172	179	7	172	180	8
140–149	14	185	195	10	173	178	5	174	188	14	170	183	13
150–	7	193	205	7	165	187	18	173	182	9	168	179	11
All	58	180	189	9	168	178	10	173	183	10	172	182	10

The table should be read as follows: Fourteen pupils with intelligence quotients between 125 and 130 had an average mental age of 175 months on the first test and an average mental age of 181 months on the second test. The gain in mental age was 6 months. They had an average educational age in nature study and science of 169

months on the first test and of 171 months on the second test. The gain was 2 months. In history and literature the average educational age was 172 months on the first test and 183 on the second test. The gain was 11 months. The average total educational age on the first test was 176 months and on the second test was 185 months. The gain was 9 months. Twenty-three pupils with intelligence quotients between 130 and 139 had an average mental age of 178 months on the first test and of 183 on the second test. The gain was 5 months. The average educational age in nature study and science was 170 months on the first test and 175 on the second test. The gain was 5 months. In history and literature the average educational age was 172 months on the first test and 179 months on the second test. The gain was 7 months. The total educational age on the first test was 172 months and on the second test 180 months. The gain was 8 months. Fourteen pupils with intelligence quotients between 140 and 149 had an average mental age of 185 months on the first test and of 195 on the second test. The gain was 10 months. The average educational age in nature study and science on the first test was 173 and on the second test was 178. The gain was 5 months. The average educational age in history and literature was 174 on the first test and 188 on the second test. The gain was 14 months. The average total educational age was 170 months on the first test and 183 on the second test. The gain was 13 months. Seven pupils with intelligence quotients of 150 or higher had an average mental age of 193 months on the first and 205 on the second test. The gain was 12 months. Their average educational age in nature study and science was 165 months on the first test and 187 on the second test. The gain was 18 months. Their average educational age in history and literature on the first test was 173 and on the second test was 182. The gain was 9 months. The average total educational age on the first test was 168 and the second 179. The gain was 11 months. For the entire group of 58 pupils the average mental age at the time of entering was 180 months. On the second test, given 5 months later, the average mental age 189 months, a gain of 9 months. The average educational age in nature study and science on the first test was 168 and on the second test 178, a gain of 10 months. In history and literature the average educational age was 173 on the first test and 183 on the second test, a gain of 10 months. The average total educational age on the first test was 172 months and on the second test 182 months, a gain of 10 months.

The differences between mental ages and educational ages at the time of entering and at the time of the second test were computed. These differences for Group I are shown in Table II.

TABLE II.—A COMPARISON OF THE DIFFERENCES BETWEEN MENTAL AGES, EDUCATIONAL AGES AND ACCOMPLISHMENT RATIOS OF 58 PUPILS (GROUP I) OF SUPERIOR MENTAL ABILITY TESTED AT ENTRANCE AND AGAIN FIVE MONTHS LATER

IQ	Test	MA	Science and nature study			History and literature			Total		
			EA	Dif.	AR	EA	Dif.	AR	EA	Dif.	AR
125–129	I	175	169	−6	..	172	−3	..	176	1	
	II	181	171	−10	..	183	2	..	185	4	
130–139	I	178	170	−8	..	172	−6	..	172	−6	
	II	183	175	−8	..	179	−4	..	180	−3	
140–149	I	185	173	−12	..	174	−11	..	170	−15	
	II	195	178	−17	..	188	−7	..	183	−12	
150–	I	193	165	−28	..	173	−20	..	168	−25	
	II	205	187	−18	..	182	−23	..	172	−33	
All	I	180	168	−12	93	173	−7	96	172	−8	96
	II	189	178	−11	94	183	−6	97	182	−7	96

This table shows that the group of pupils with intelligence quotients between 125 and 129 and an average mental age of 175 months initially, had an average educational age in nature study and science of 169, which is 6 months lower than from their mental age average. The average educational age in history and literature was 172 months, 3 months lower than the mental age. The average total educational age was 176, 1 month higher than the mental age. During the 5 months between tests the increase in achievement in nature study and science did not keep pace with the increase in mental age. This is true also for the group with intelliegence quotients between 140 and 149. The group between 130 and 139 in intelligence increased in knowledge of nature study and science at the same rate as they increased in mental age, while the highest group, those with intelligence quotients of 150 or higher, lessened the difference between their ability to do and their

accomplishment by 10 months. Considering the difference between the mental age and educational age in nature study and science for the whole group, we find an average difference of 12 months on the first test and an average difference of 11 months on the second test. Expressed in terms of the ratio of actual accomplishment to possibility of accomplishment, an improvement had been made from ninety-three per cent at the beginning to ninety-four per cent at the second test 5 months later.

Results similar to these are shown in studying the data for history and literature. The initial difference between mental age and educational age was less, consequently the probability of improvement was less. An average difference of 7 months for the entire group at the beginning was decreased to 6 months, changing the achievement ratio from ninety-six to ninety-seven per cent. In total educational age the initial achievement remained constant.

The gain in mental age and in educational ages for Group II, who who were tested at intervals one school year apart, are shown in Table III.

TABLE III.—GAINS IN MENTAL AGES AND IN EDUCATIONAL AGES OF 94 PUPILS (GROUP II) OF SUPERIOR MENTAL ABILITY MEASURED BY TESTS GIVEN ONE YEAR APART

IQ	N	MA			EA in nature study and science			EA in history and literature			Total EA		
		Test I	Test II	Gain	Test I	Test II	Gain	Test I	Test II	Gain	Test I	Test II	Gain
125–129	37	158	173	15	151	172	21	159	176	17	160	179	19
130–139	31	165	181	16	158	181	23	165	183	18	161	182	21
140–149	17	177	191	14	161	183	22	165	177	12	164	182	18
150–	14	192	208	16	175	195	20	178	192	14	179	195	16
All	94	168	184	16	159	180	21	165	180	15	164	181	17

This table shows a comparatively consistent gain that is beyond normal expectancy. For the entire group of 94 pupils the average increase in educational age in nature study and science was 21 months, in history and literature 15 months, and in total educational age 17 months.

The differences between the mental ages and the educational ages for this group are shown in Table IV.

A comparison of Table III and Table IV brings out an interesting truth. The growth in knowledge, as measured by the acquisition of facts, had been rapid beyond expectation, but mental age also had increased. The relation between possibility of achievement and actual achievement had approached perfection but had not reached it.

TABLE IV.—A COMPARISON OF THE DIFFERENCES BETWEEN MENTAL AGES, EDUCATIONAL AGES, AND ACCOMPLISHMENT RATIOS OF 94 PUPILS (GROUP II) OF SUPERIOR MENTAL ABILITY TESTED AT ENTRANCE AND ONE YEAR LATER

IQ	Test	MA	Nature study and science			History and literature			Total		
			EA	Dif.	AR	EA	Dif.	AR	EA	Dif.	AR
125–129	I	158	151	−7	..	159	1	..	160	2	
	II	173	172	−1	..	176	3	..	179	6	
130–139	I	165	158	−7	..	165	0	..	161	−4	
	II	181	181	0	..	183	2	..	182	1	
140–149	I	177	161	−16	..	165	−12	..	164	−13	
	II	191	183	−8	..	177	−14	..	182	−9	
150–	I	192	175	−17	..	178	−14	..	179	−13	
	II	208	195	−13	..	192	−16	..	195	−13	
All	I	168	159	−9	95	165	−3	98	164	−4	98
	II	184	180	−4	98	180	−4	98	181	−3	98

The pupils were in the position of a man who attempts to acquire capital at the time he is paying off indebtness. However we find the average accomplishment ratio was ninety-eight per cent.

Seventy-two of the pupils of this group were the children of professional people with the opportunities for good reading that are usually found in the homes of educated people. Almost the opposite was true of the next group, a majority of whom were in a room that had been established in a neighborhood peopled by workers in a tire factory. The opportunities for reading furnished by the homes were most meager, and the library was built up very slowly. Perhaps these conditions reflected in the differences in the data.

Table V and Table VI present the data for Group III. An interval of one and a half years elapsed between the two tests.

TABLE V.—GAINS IN MENTAL AGES AND EDUCATIONAL AGES OF 45 PUPILS OF SUPERIOR MENTAL ABILITY MEASURED BY TESTS GIVEN ONE AND ONE-HALF YEARS APART

IQ	N	MA			EA in nature study			EA in history			Total EA		
		Test I	Test II	Gain	Test I	Test II	Gain	Test I	Test II	Gain	Test I	Test II	Gain
125–129	19	156	175	19	142	163	21	137	165	28	151	168	17
130–139	11	168	193	25	158	172	14	152	177	25	163	182	19
140–149	6	175	197	22	156	178	22	155	176	21	157	186	29
150–	9	181	209	28	166	190	24	177	195	18	173	197	24
All	45	167	189	22	152	173	21	151	175	24	159	180	21

TABLE VI.—A COMPARISON OF THE DIFFERENCE BETWEEN MENTAL AGES, EDUCATIONAL AGES, AND ACCOMPLISHMENT RATIOS OF 45 PUPILS (GROUP III) OF SUPERIOR MENTAL ABILITY TESTED AT ENTRANCE AND ONE AND ONE-HALF SCHOOL YEARS LATER

IQ	Test	MA	Nature study and science			History and literature			Total		
			EA	Dif.	AR	EA	Dif.	AR	EA	Dif.	RA
125–129	I	156	142	−14	..	137	−19	..	151	−5	
	II	175	163	−12	..	165	−10	..	168	−7	
130–139	I	168	158	−10	..	152	−16	..	163	−5	
	II	193	173	−21	..	177	−16	..	182	−11	
140–149	I	175	156	−19	..	155	−20	..	157	−18	
	II	197	178	−19	..	176	−21	..	186	−11	
150–	I	181	166	−25	..	177	−4	..	173	−8	
	II	209	190	−19	..	195	−14	..	197	−12	
All	I	167	152	−15	91	151	−16	90	159	−8	95
	II	189	173	−16	91	175	−14	92	180	−9	95

The growth in achievement was satisfactory when measured from the starting point of the mental age at the initial test, but it was disappointing when considered in relation to the increased mental age. The slight increase in accomplishment ratio leaves it still approximately ten points too low. It may be assumed that the

ninety-five per cent achievement in total education was due to the two tests in arithmetic that are a part of the Stanford Achievement Test battery.

The Stanford Achievement Tests were published, at least they were available for the purpose of testing Group IV after the pupils had been subjected to the reading program about a year. For that reason the educational status of the group must be considered as already somewhat higher relatively than that of the three other groups.

Table VII, shows the gain for Group IV during a period of two years.

TABLE VII.—GAINS IN MENTAL AGES AND IN EDUCATIONAL AGES OF 30 PUPILS (GROUP IV) OF SUPERIOR MENTAL ABILITY MEASURED BY TESTS GIVEN TWO YEARS APART

IQ	N	MA			EA in nature study			EA in history			Total EA		
		Test I	Test II	Gain	Test I	Test II	Gain	Test I	Test II	Gain	Test I	Test II	Gain
125–129	11	146	179	33	148	183	34	149	188	39	152	190	38
130–139	10	163	194	31	155	191	36	161	191	30	159	198	39
140–149	3	176	208	32	163	207	44	168	196	28	163	198	35
150–	6	182	201	19	165	218	53	169	207	38	170	208	38
All	30	162	191	29	156	195	39	159	194	35	157	197	40

It will be noted that in all mental levels the average educational gain exceeded the mental gain. It will be noted also that, contrary to expectation, the gain in the upper levels was not less than that in the lower levels. Table VIII, which follows, indicates the reason for this more clearly.

The eleven pupils whose intelligence quotients fall within the 125–129 interval were achieving at a level above expectancy at the time the first tests were given, probably because they had been following the special course of study for a year. There was therefore little room for further improvement. However, they made the most of that little and raised their level of achievement still higher above the level expected of them.

The six pupils with intelligence quotients above 150 showed achievement on the first test in agreement with the published findings of Stanford University. That is, their accomplishment in nature study

TABLE VIII.—A COMPARISON OF THE DIFFERENCES BETWEEN MENTAL AGES, EDUCATIONAL AGES, AND ACCOMPLISHMENT RATIOS OF 30 PUPILS (GROUP IV) OF SUPERIOR MENTAL ABILITY TESTED AT INTERVALS OF TWO YEARS

IQ	Test	MA	Nature study and science			History and literature			Total		
			EA	Dif.	AR	EA	Dif.	AR	EA	Dif.	AR
125–129	I	146	149	3	..	149	3	..	152	6	
	II	179	183	4	..	188	9	..	190	11	
130–139	I	163	155	−8	..	161	−2	..	159	−4	
	II	194	191	−3	..	191	3	..	198	4	
140–149	I	176	163	−13	..	168	−8	..	163	−13	
	II	208	207	−1	..	196	−12	..	198	−10	
150–	I	182	165	−17	..	169	−13	..	170	−12	
	II	201	218	17	..	207	8	..	208	7	
All	I	162	156	−6	96	159	−3	98	157	−5	97
	II	191	195	4	102	194	3	102	197	6	103

and science was ninety-one per cent, in history and literature ninety-three per cent, and in total education ninety-three per cent of normal expectancy. At the end of the two years they had surpassed normal expectation. The table shows that, for the thirty pupils constituting this two year group, the achievement ratio in each of the subjects was slightly more than one hundred per cent at the end of the two years.

CONCLUSIONS

It is discreet to submit the statistical findings and leave the thoughtful reader to draw his own conclusions. However, to the writer the following conclusions are indicated:

1. The course of study and the methods in general use in the elementary schools do not produce in children of superior mental ability the level of achievement of which they are capable.

2. A higher level of achievement is reached by means of more varied and more extensive reading of an interesting character if some use is made of the material read.

3. That time is needed to overcome the acquired inertia is apparent in the fact that improvement is slow and gradual.

4. The study does not indicate that among children of superior mental ability the degree of improvement or the rate of improvement is dependent upon the degree of intelligence.

5. It is not safe to assume that gain in the acquisition of information is adequately measured by the Stanford Achievement Tests. Further experimentation of the character of this study with other tests are needed.

SYSTEMATIC DIFFERENCES IN THE VARIOUS PARTS OF THE HERRING REVISION OF THE BINET-SIMON TEST WHEN APPLIED TO NORMAL DULL ADULTS

H. H. REMMERS

Purdue University

In connection with another study, the writer recently had occasion to apply the Herring Revision of the Binet Tests to one hundred one freshmen men selected from the lower fourth of the total freshmen population at Purdue University on the basis of certain group tests. An analysis of the performance of these selected freshmen revealed what appear to be significant differences in the various parts[1] of the revision. Tables I, II, and III give the pertinent data (slide-rule calculations). All of the students represented in the tabulations were tested by the writer himself.

As appears in Table I, the mental age from each part was tabulated for all subjects. Table II gives an analysis of the differences observed. The last column of Table II gives the "index of significance" (the ratio of the observed difference to its PE). It will be observed that the differences for the following pairs of parts of the tests are statistically significant:[2]

A-B	B-E
A-C	C-D
B-C	C-E
B-D	

Table III gives the observed differences in the standard deviations for the various parts of Herring revision. With the exceptions of the differences for B and C, B and E, and C and D, all the differences are significant.

---

[1] See: "Herring Revision of the Binet-Simon Tests." Examination Manual, Form A. New York: Yonkers-on-Hudson, World Book Co., 1923.

[2] "Statistically significant" is defined as an index of significance greater than three. This indicates that the chances are at least 98 in 100 that the observed difference is a true difference and will not disappear with further sampling (See Garrett, H. E.: "Statistics in Psychology and Education." P. 135). With an index of four the chances are 99.7 in 100. The generalizations based on the data in this paper are very conservative since the $PE_{diff.}$ was calculated from the formula for uncorrelated measures. There is, of course, very substantial correlation.

TABLE I.—HERRING REVISION OF THE BINET-SIMON TEST MENTAL AGES

Part	A	B	C	D	E
215–219	1				
210–214	1				
205–209	..	..	..	..	1
200–204	3	..	..	..	2
195–199	7	1	..	1	2
190–194	8	..	..	4	8
185–189	11	4	4	14	12
180–184	..	2	14	16	16
175–179	11	13	10	15	16
170–174	11	17	20	16	14
165–169	12	7	20	11	11
160–164	13	22	13	11	9
155–159	12	11	7	7	6
150–154	..	8	6	2	
145–149	3	8	2		
140–144	6	4	2	1	1
135–139	1	1	1	1	2
130–134	..	..	1	2	1
125–129	..	1	1		
120–124	1	2			
N	101	101	101	101	101
Mean	172.100	163.140	167.700	172.450	175.650
σ	17.550	13.200	10.920	10.440	13.400
PE_σ	0.833	0.626	0.518	0.496	0.636
PE_M	1.190	0.890	0.733	0.700	0.899

TABLE II.—DIFFERENCES OF MEANS OF THE VARIOUS PARTS IN THE HERRING REVISION

	Difference	$PE_{diff.}$	$\frac{Diff.}{PE_{diff.}}$	Chances in 100
A-B	8.960	1.484	6.05	100
A-C	4.400	1.400	3.15	98
A-D	0.350	1.380	0.25	57
A-E	3.550	1.430	2.48	95
B-C	4.560	1.155	3.94	100
B-D	9.310	1.134	6.95	100
B-E	12.510	1.263	9.92	100
C-D	4.750	1.015	4.68	100
C-E	7.950	1.160	6.85	100
D-E	3.200	1.168	2.74	97

TABLE III.—DIFFERENCES OF STANDARD DEVIATIONS OF THE VARIOUS PARTS OF THE HERRING REVISION

	Difference	$PE_{diff.}$	$\frac{Diff.}{PE_{diff.}}$	Chances in 100
A-B	4.35	1.044	4.17	100
A-C	6.63	0.981	6.76	100
A-D	7.11	0.970	7.33	100
A-E	4.15	1.048	3.96	100
B-C	2.28	0.815	2.80	97
B-D	2.76	0.801	3.44	99
B-E	0.20	0.894	0.22	56
C-D	0.48	0.717	0.67	67
C-E	2.48	0.820	3.02	98
D-E	2.96	0.807	3.66	99

How are these differences to be accounted for? One possibility, of course, is that systematic differences were introduced by the administration of the tests. It is barely possible that the writer departed from standard procedure in a consistent fashion in the application or scoring (or both) for some items or for some parts of the tests more than for others. This seems unlikely in view of the fact that the directions are quite specific and the scoring relatively highly objective, and also in view of the fact that this was not the writer's first experience in the use of the Herring Revision. He had previously used it with something like forty cases of children ranging from chronological age six to eighteen.

A more plausible explanation would seem to lie in the direction of the assumption that the differences represent not errors of method, but genuine differences in mental functions for the group in question as compared with the unselected sampling represented by the Herring Revision norms.

Before going on with a consideration of this hypothesis, it will be well to recall that the data presented above are concerned with a group a large proportion of which may be classified as "normal dull" subjects. They represent the lower end of the distribution of ability of college freshmen on the basis of the median percentile score on a battery of six quasi-intellectual and educational group tests.[1] Tables

[1] The tests are as follows: National Research Council on Education Test (1926 ed.), Iowa Placement Examinations in Chemistry Aptitude, Mathematics Aptitude, Physics Aptitude, English Aptitude and English Training. For a detailed

TABLE IV.—HERRING REVISION OF THE BINET-SIMON TEST IQ's

IQ	A	B	C	D	E
110–112	1				
107–109	4	..	..	..	1
104–106	..	..	..	..	2
101–103	7	1	..	1	4
98–100	8	1	2	9	11
95– 97	11	5	6	19	20
92– 94	11	13	19	18	16
89– 91	11	11	21	19	16
86– 88	13	22	20	10	12
83– 85	12	13	13	12	10
80– 82	12	11	7	7	5
77– 79	9	15	8	2	
74– 76	..	2	1	..	1
71– 73	..	3	1	2	2
68– 70	1	2	2	1	1
65– 67	..	1	1	1	
62– 64	1	1			
N	101	101	101	101	101
Mean	89.733	85.950	86.763	91.695	91.484
σ	8.820	6.225	6.735	5.790	7.248
PE_M	0.592	0.417	0.452	0.388	0.486

TABLE V.—DIFFERENCES OF THE MEANS IN THE VARIOUS PARTS OF THE HERRING REVISION

Part	Difference	$PE_{diff.}$	$\frac{Diff.}{PE_{diff.}}$	Chances in 100
A-B	3.883	0.724	5.37	100
A-C	2.970	0.745	4.02	100
A-D	1.962	0.708	2.76	97
A-E	1.751	0.766	2.78	97
B-C	0.813	0.615	1.38	83
B-D	5.745	0.570	10.01	100
B-E	5.534	0.641	8.63	100
C-D	4.932	0.596	8.22	100
C-E	4.721	0.664	7.12	100
D-E	0.211	0.622	0.34	59

description of this group of students see H. H. Remmers: "A Diagnostic and Remedial Study of Potentially and Actually Failing Students at Purdue University." Bulletin of Purdue University, Vol. XXVII, No. 14, Educational Reference Circular I, Apr., 1927, pp. 88.

IV and V have been prepared to show the IQ achievement and the differences of means.

The hypothesis previously advanced that the observed differences represent actual differences in mental functions or processes receives support from previous studies. Cunningham in a recent study[1] summarizes the literature and reports an original investigation on the comparison of normal with inferior intelligence. A paragraph from his summary of the literature is apposite:

"So far as comparisons of normal with inferior minds are concerned, it does appear that there are genuine differences in their performance on different kinds of tests. There does, also, seem to be some strand of similarity running through the tests which various investigators have selected as being the most discriminative. But in all our comparisons of those of equivalent mental ages but different degree of ability (different IQ's) it must be remembered that those of poorer ability have lived longer in the world and that it is possible to claim that any differences in performance are accountable for by this fact."[2]

In his comparison of the mental functions of a fairly normal sampling of children of chronological age 2.5 to 5.5 years with adult imbeciles of the same mental age (CAVD Test) he shows that certain types of test material are responsible for systematic differences in performance and also that there seems to be a certain mental age level (MA = 4 years) at which "length of experience and ability to profit by it counterbalance each other."[3] In concluding his analysis of differences he states: "If any one general inference is possible, it would seem to be that the tests involving an appreciation of relative size and those involving an appreciation of relationship of given to incomplete parts are characteristically easier for the group representing normal intelligence."[4]

No very direct inferences from Cunningham's study of normal children and adult imbeciles can, of course, be made in relation to the data on normal dull adults presented earlier in this paper. If his tentative conclusion that there exists a definite relation between length of experience and ability to profit from it is sound, however, then it has a bearing on these data. It would seem a plausible infer-

---

[1] Cunningham, K. S.: "The Measurement of Early Levels of Intelligence." Teachers College Contributions to Education, No. 259, 1927, pp. 35–67.

[2] *Op. cit.*, p. 44.

[3] *Op. cit.*, p. 48.

[4] *Op. cit.*, p. 62.

ence that the test items in the Herring Revision vary for different parts of the tests in revealing such a relationship.

An item-by-item analysis of the Herring Revision, especially if comparisons with normal adults were available, would very likely throw additional light on the problem involved in this study. One's psychological preconceptions—whether one conceives of mental functions as bonds, as Gestalten, as conditioned responses, or in terms of the new faculty psychology of Professor Spearman[1]—would, of course, incline one to view the data presented in one or another light, depending upon which particular "idols of the den" one happened to worship.

[1] See his "Abilities of Man." The use of the term "faculty psychology" is not meant to imply any gibe at Professor Spearman's brilliant and carefully organized researches. It has reference merely to the fact that his analysis appears to be leading to a number of stable factors or functions (*g*, *s*, and *c*) of mind analogous to the faculties of a psychology not yet as moribund, as he points out, as the textbooks would have us believe.

THE RELATION OF LEFT-HANDEDNESS TO PSYCHO-NEUROTIC TRAITS

R. G. WETMORE AND G. H. ESTABROOKS

Colgate University

The following data were collected from the student body with a view to discovering whether there was any relation between left-handedness and psycho-neurotic traits. The measure of neurotic tendencies was obtained from Laird's "Personal Inventory B2," while the data in left-handedness were collected by a canvass of the student body. This canvass yielded us sixty-two students who were either totally left-handed, ambidextrous, or who had merely a tendency towards using the left hand.

Theoretically, if neurotic symptoms are in any way connected with the use of the left hand one would expect to find them most pronounced where the individual had only partically changed over; in other words, was ambidextrous. But when we break our groups up into its various sub-groups we obtain the following:

Group	Number of individuals	Mean score
Totally left-handed	7	21.426
Ambidextrous	23	19.173
Tendency towards using left hand	32	19.671
Total	62	19.675

The number of individuals contained in the first group makes this average totally unreliable. On the other hand the ambidextrous group is actually the lowest of all three so far as psycho-neurotic traits are concerned. This difference is, however, quite unreliable since there is a very large probable error to the averages, which in the case of this ambidextrous group amounts to PE = 4.3. The sections within the left-handed group are obviously too small for us to place any great reliance of differences found here, even if they did exist.

In order to give some basis of comparison between the total left-handed groups and the average right-handed individual, we selected sixty-two men in alphabetical order from the class of '30. These men were right-handed and probably represent a fair random sample

of college boys in respect to neurotic traits. The results were absolutely negative.

Group	Number of individuals	Mean score
Total left-handed group	62	19.675
Right-handed group	62	19.531

Finally we worked out the Bi-Serial *r* on these two groups as described in the "Handbook of Mathematical Statistics" by Reitz. Again results were negative, as might have been expected, the Bi-Serial *r* being .0155.

Our results would thus seem to indicate that in as far as we can judge from the measure of psycho-neurotic traits as yielded by Laird's "Personal Inventory B2" there is no relation between left-handedness and the neurotic constitution. This statement applies, of course, only to this particular study.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY JAMES E. MENDENHALL

Research Associate, Institute of Educational Research, Teachers College
Columbia University

Exceptional Children

Children of High Intelligence. A Follow-up Enquiry. James F. Duff. The British Journal of Psychology, April, 1929, 413–438. A follow-up four years after testing of seventy-three children over 136 IQ. The gifted group was superior to the control (*N*, 28, average IQ of 100) in academic success, teacher's trait ratings, occupations of parents, etc.,

Manual Motor Correlation in Superior Children. Frieda A. Kiefer. The Journal of Applied Psychology, Aug., 1929, 357–371. In the tensiometer, tapping, steadiness, tracing, strength of grip, rating of manual abilities, the differences between superior and average groups of children were negligible.

Follow-up Investigation of Five Hundred Children Who Previously Attended Classes for the Physically Handicapped. Jessie M. Lightfoot. The Psychological Clinic, March-April 1929, 1–16.

Characteristics of Problem Children Based on Judgments of Teachers. W. E. McClure. The Journal of Juvenile Research, April, 1927, 124–140. A report of the characteristics of five hundred thirty-three children Grades I to VIII showed (1) that the average chronological age of problem children was from 4.4 to 21.2 months greater than that of normal children, and (2) that the mean IQ of problem children was from 85 to 99.

A Study of Alameda County Delinquent Boys, with Special Emphasis upon the Group Coming from Broken Homes. Sarah B. Crosby. The Journal of Juvenile Research, July, 1929, 220–230. A greater incidence of delinquency appears among boys with broken homes than among boys with complete homes.

Problems Presented and Results of Treatment in One Hundred Fifty Cases Seen at the Habit Clinic for Pre-school Children in Boston. Sybil Foster and Dorothy Stebbins. Mental Hygiene, July, 1929, 529–541. This report includes an analysis of one hundred fifty cases according to habits and improvement in those habits.

Attitudes and Educational Disabilities. Phyllis Blanchard. Mental Hygiene, July, 1929, 550–563. A report of five case studies of disabilities in reading or arithmetic, arising from the attitudes of parents.

A Sociological Case Study of a Foster Child. Walter C. Reckless. The Journal of Educational Sociology, June, 1929, 567–584. An interesting, intensive study of a child, seven years of age, an orphan from the mountains of Tennessee.

How a Special Disability in Spelling Was Diagnosed and Corrected. William F. Book. The Journal of Applied Psychology, Aug., 1929, 378–393. An interest-

ing report of a course of training in spelling given to a sixth grade boy, Stanford-Binet IQ of 120.

Child Psychology

Child Development Abstracts and Bibliography. Committee on Child Development, National Research Council, Washington, D. C., Aug., 1929.

Maturation and Infant Behavior. Arnold Gesell. Psychological Review, July, 1929, 307–319. An interesting description of experiments at the Yale Psycho-clinic and their implications.

The Speech Development of Four Infants under Two Years of Age. Katherine Foulke and Sarah M. Stinchfield. Journal of Genetic Psychology, March, 1929, 140–171.

Some Suggestive Results Regarding Sex Differences in Attitude toward School Work. Paul A. Witty and Harvey C. Lehman. Education, April, 1929, 449–458. Tables show occupational choices of children of various ages for a career in education.

Punishments Recommended for School Offenses: A Reply. Frances Dummer and Carleton Washburne. The Elementary School Journal, June, 1929, 774–786. Interesting accounts of the way in which several discipline cases were handled.

Educational and Vocational Guidance

The Prediction of Scholastic Success by Means of Classification Examinations. Philip M. Condit. Journal of Education Research, May, 1929, 331–335. With five hundred fifty-nine entering students the Thurstone test correlated with scholarship, .45; and the achievement test with scholarship, .50.

The Effectiveness of Certain Requirements in the Selective Admission of College Students. Edward F. Potthoff. The School Review, September, 1929, 519–530. Tables show for average high school marks comparisons between successful students and unsuccessful students.

High School Subjects as Conditioners of College Success: Implications and Theories Concerning Mental Factors and Faculties. Herbert Sorenson. Journal of Educational Research, April, 1929, 237–254. The writer finds that "Latin differentiates according to college success more markedly than any other subject," and that "success in Latin marks a student as being a good one."

The Relative Rank in High School and in the First Two Years of the University. J. Harold Goldthorpe. School and Society, July 27, 1929, 130–134. A study of one hundred thirty-six entering college freshmen showed that "there appears to be a substantial relationship between rank in high school and standing in the first two years in university."

Predicting Success in First Semester College Courses in Physical Science. M. Eustace Broom and J. W. Lawson. School Science and Mathematics, June, 1929, 623–626. The study reports these correlations (N, 185): Average grade points per unit of physical science in college and (1) number of units in high school physical science, .141; (2) average grade points per unit of physical science in high school, .313; (3) Thorndike total equivalent score, .355; (4) Thorndike reading test decile ranks, .209; and (5) Thorndike trade information decile ranks, .123.

The Exploratory-guidance Value of General Language. Walter Kaulfers. California Quarterly of Secondary Education, June, 1929, 307–312. The correlations between general language grades and foreign language grades were .486 for seventy-eight boys and .369 for one hundred eight girls.

Elusive Factors Tending to Reduce Correlations between Intelligence Test Ranks and College Grades. Frank S. Freeman. School and Society, June, 14, 1929, 784–786. On the questionnaire superior students in intelligence gave the following explanations for poor grades: lack of interest, poor habits of study, etc. Students inferior in intelligence test gave the following explanation for their low rating and high scholarship: newness of conditions and nervousness during first days, attached no importance to test, etc.

Remediation of Teachers College Freshmen. Walter Scribner Guiler. School and Society, Aug. 17, 1929, 242–244. College freshmen (*N*, 108 to 341) were (1) below the twelfth grade norms in the following tests: Monroe Silent Reading, thirty-seven per cent; Guiler Spelling, forty-two per cent; Hudelson Composition, eighty-two per cent; Guiler-Henry Punctuation, seventy-nine per cent; Gregory American History, eighty-three per cent; and (2) below the eighth grade norms in the following: Ayers Handwriting, eighteen per cent; Monroe Arithmetical Computation, twenty-three per cent; Buckingham Problem-Solving, sixty-nine per cent; Buckingham U. S. Geography, twenty per cent. Remedial instruction in spelling and punctuation netted large gains in test scores.

What a Freshman Wants to Know. Lester Raines. School and Society, May 4, 1929, 577–578. A questionnaire (*N*, 125) showed the following points to be interesting to students: Fraternities and sororities, sixteen per cent; Classification and teaching staff, thirteen per cent, etc.

Prognosis of Musical Achievement. Hazel M. Stanton. Studies in Psychology, University of Rochester, Vol. I, No. 4, 1929. A validation over a period of years of the Seashore tests administered to three hundred fifty-one entering students showed a positive association between test scores and scholarship, honors, survival, etc.

A Study of Parent, Teacher and Student Cooperation in High School. School and Society, Apr. 27, 1929, 553–556. Returns from questionnaires submitted to five hundred forty-eight students, twenty-three teachers, and two hundred fifty-three parents.

Nonschool Educational Agencies. F. R. Clow. The Journal of Educational Sociology, June, 1929, 593–598. A study of 2000 children in 1915 and 1916 showed from Grades V to high school (1) an increasing per cent doing school work at home, especially in high school; (2) an increasing per cent doing other reading (except in high school); (3) an increasing per cent doing other studying (except in high school); (4) a constant per cent doing useful work at home; (5) an increasing per cent working elsewhere for pay; (6) an increasing per cent attending the theatre or movies; (7) a decreasing per cent retiring before ten o'clock.

The Avocations of One Hundred College Freshmen. Stuart M. Stoke and W. F. Cline. The Journal of Applied Psychology, June, 1929, 257–265. The study shows that (1) sedentary forms of recreation are most popular; (2) the more intelligent play twenty-five per cent more time than the less intelligent; (3) few differences between bright and dull in types of occupations (excluding reading), etc.

The Measurement of Interest Differences between Students of Engineering and of Agriculture. H. H. Remmers. The Journal of Applied Psychology, April, 1929, 105–119. With two samples of one hundred students each, the interest check list differentiated engineering and agricultural students.

Interest of Engineers. A Basis for Vocational Guidance. Edward K. Strong, Jr. The Personnel Journal, April, 1929, 441–454. Data secured from 3920 men in thirty occupations were used to validate the items of the test.

A Study of Vocational Interests of California High School Students Based on a Survey of Twelve Rural Schools. J. H. Bedford. California Quarterly of Secondary Education, June, 1929, 277–295. This article includes statistics from a survey of the follow-up records of 1478 students.

Six Year's Progress—The Present Status of Counseling in the High Schools of California. California Quarterly of Secondary Education, June, 1929, 296–306. This study reports guidance activities in table form and suggests steps for the future.

Some Problems in Junior Placement. Clare Lewis. The Personnel Journal, August, 1929, 125–137. An interesting report of the guidance activities in a public school system.

The Correspondence of School Achievement and Industrial Efficiency with Mental Age as Obtained by the Stanford-Binet. Catherine E. Chipman. The Psychological Clinic, March-April 1929, 21–28. Intelligence rating seemed to have a negligible association with "good work on the job."

Technopsychology in a Swiss Industry. Leon Walther. The Personnel Journal, June, 1929, 1–18. An intensive study of two hundred thirty women workers in a factory, using job analyses and test techniques.

The Human Factor in Substation Operation. Morris S. Viteles. The Personnel Journal, August, 1929, 81–113. Paper and pencil tests and performance tests are described.

Measuring Judgment and Resourcefulness. L. J. O'Rourke. The Personnel Journal, April, 1929, 427–440. The article describes the use of an oral problem for interview examination to determine fitness of an applicant for the prohibition enforcement service.

Business Girls. A Study of Their Interests and Problems. Ruth Shonle Cavan. The Religious Education Association, Chicago, June, 1929. Evidence concerning five hundred ninety-one girls is presented under these headings: Office workers, interests and problems, home and family life, intellectual abilities and interests, friends, recreation, vocational problems, money, church, personality, etc.

List of References on Vocational Guidance. Martha R. McCabe. Department of the Interior, Bureau of Education, Washington, D. C., May, 1929.

What the Chicago Doctors of Philosophy Are Doing. M. O. Wilson. School and Society, June 22, 1929, 815–818.

Social Problems

The Measurement of Success in Marriage and in Parenthood. Laura Chassell Toops. Teachers College Record, March, 1929, 579–588. A preliminary outline of steps to investigate this problem.

Decreases in Intelligence with Increase in Age among Inmates of Penal Institutions. George Rex Mursell. The Journal of Juvenile Research, July, 1929,

197–203. With large numbers of adults, there was a decrease in IQ from one chronological age group to the next.

Recent Statistics of Alcoholic Mental Disease. Horatio M. Pollock and Frederick W. Brown. Mental Hygiene, July, 1929, 591–614.

Emotion and the Incidence of Disease, the Influence of the Number of Diseases, and of the Age at Which They Occur. George M. Stratton. The Psychological Review, May, 1929, 242–253. The writer concluded from a study of 2600 cases that those who had been subject to disease responded more intensely to anger and fear situations. Diseases occurring between eleven and fifteen years of age were most important in fear reactions; those before six, in anger reactions.

Philosophy of Education

The Most Outstanding Next Steps for Curriculum Makers in the United States. Boyd H. Bode. Teachers College Record, December, 1928, 179–191.

The Relations of Philosophy and Science in the Study of Education. William H. Kilpatrick, School and Society, July 13, 1929, 39–48.

Criteria for Judging a Philosophy of Education. George S. Counts. School and Society, July 27, 1929, 103–107.

The Contribution of Science to Education. Frank N. Freeman. School and Society, July 27, 1929, 107–112.

Miscellaneous

The Seashore Measures of Musical Talent. M. E. Broom. School and Society, Aug. 24, 1929, 274–275. With eighty-two college upperclassmen and one hundred two ninth grade pupils these correlations were secured: (1) *Pitch* and (*a*) Intensity, .39 and .92; (*b*) time, .29 and .90; (*c*) consonance, .28 and .89; (*d*) memory, .41 and .88; (*e*) rhythm, .34 and .90. (2) *Intensity* and (*a*) time, .17 and .96; (*b*) consonance, .30 and .87; (*c*) memory, .35 and .89; (*d*) rhythm, .08 and .86. (3) *Time* and (*a*) consonance, .31 and .86; (*b*) memory, .24 and .88; (*c*) rhythm, .32 and .83. (4) *Consonance* and (*a*) memory, .36 and .94; (*b*) rhythm, .18 and .89. (5) *Memory* and rhythm, .38 and .90. The writer concludes that the tests measure different aspects of the general capacities at various age levels.

A Year of Research—1927. Goodwin B. Watson and Delia H. Biddle. The Religious Education Association, Chicago, July, 1929. An annotated summary of studies bearing upon the program of religious, educational and social agencies.

Great Teachers and Some Methods of Producing Them. Robert L. Kelly and Ruth E. Anderson. Journal of Educational Research, June, 1929, 22–30. A tabulation of returns from 162 college executives to show the academic training, the term of service, the subjects taught, the qualities and attitudes of great teachers; and the methods of encouraging best students for college teaching.

Contributions to Educational Periodicals. E. A. Collins. Peabody Journal of Education, Mar., 1929, 278–281. An analysis of publications from 1920–1926 showed that university professors contributed 45 per cent of the articles.

Who Have Contributed Most to the Educational Measurement Movement? C. W. Odell. School and Society, June 8, 1929, 751–754. Results of an analysis of 20 recent publications showed a rank order of leaders in educational measurement.

A Critical Review of Book Reviews. Walter S. Monroe and Mabel R. Hull. School and Society, Apr. 20, 1929, 521–526. The writers analyse 548 reviews by

frequency of books reviewed by magazines, by descriptive items in reviews, and by evaluation items in reviews. "Book reviews may be criticized upon at least two grounds: they are inadequate and they are non-critical."

Studies of Typographical Factors Influencing Speed of Reading. Donald G. Patterson and Miles A. Tinker. The Journal of Applied Psychology, (II) Apr., 1929, 120–130; (III) June, 1929, 205–219. With 320 college students, 10 point type proved optimal for an 80 mm. length of line.

Parent Education, 1926–1928. Ellen C. Lombard. Department of the Interior, Bureau of Education, Washington, D. C., 1929.

New Tests

Columbia Research Bureau Chemistry Test. Eric R. Jette, Samuel R. Powers, and Ben D. Wood. World Book Company, Yonkers, N. Y., 1929.

Elwell-Fowlkes Bookkeeping Test. World Book Company, Yonkers, N. Y., 1929.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

Contemporary Education; Its Principles and Practice, by Paul Klapper. New York: D. Appleton and Co. Pp. XXV + 650.

In this volume Professor Klapper has undertaken "to coordinate the accepted principles of social and psychological studies and to indicate their significance for education today." The five parts of the book under the main captions of the Meaning and Function of Education, Education as Physical Adjustment, Education as Social Adjustment (two parts), Education as Economic Adjustment, and Education as Mental Adjustment deal almost encyclopedically with practically every problem and question with which the beginner in education is likely to be confronted—aims, curriculum, methods, educational psychology, and vocational education. While the scope of the work is extensive, it is characterized by remarkable simplicity and lucidity of style. The concrete examples drawn from practical schoolroom situations contribute in no small degree to a clarification of the subject for the beginner. Throughout the book Professor Klapper keeps his feet on the ground, discusses every aspect of the problem and, because he has his eye on the practical purpose that he has set himself, refrains from too early exposure of the student to the conflicts injected into education by frontier thinkers and progressive schools. The suggested readings and questions for discussion in each chapter add to the usefulness of the work for classroom instruction with the groups for which it is intended. I. L. KANDEL.

International Institute, Teachers College, Columbia University.

The Theory of Identical Elements, by Pedro Temesis Orata. The Ohio State University Press, 1928. Pp. XI + 204.

Another lance, neither sharp enough nor strong enough to do any damage, has been hurled against the Thorndikian shield.

As early as his Preface Dr. Orata involves himself in a fundamental difficulty when he makes as one of his main points of attack the incompatability of Thorndike's psychology "with the democratic ideal," and sets up as his principal aim the attainment of a "theory of transfer of training which makes for the realization of the democratic ideal which the school is expected to promote in the pupils." It is difficult to see just what a political concept has to do with a scientist's unprejudiced approach to a problem. Orata's book is another example of the pernicious influence of philosophy on psychology. Each has its place, and for the scientific worker to mix them is to arrive at muddled conclusions.

The writer does give numerous excellent summaries of experiments that have been made in the transfer field. From these he attempts to show the weakness of the theory of identical elements and the adequacy of the generalization point of view, although just how the two differ, if at all, is not made clear.

As would be expected there is an attack on Thorndike's inability to explain exactly how transfer takes place. Probably that is the best point that Dr. Orata makes. However, he himself becomes very hazy when he offers, in the second half of his book, a reinterpretation of the problem. There is much argument, with attendant confusion, concerning "training in thinking" (faculty psychology has many more than nine lives) as opposed to mechanical habit formation. The author asserts that experimental evidence is lacking for Thorndike's point of view, but it is not apparent as to how and where his own contentions are as well buttressed.

The book then, is not a clear, scientific attack on the weaknesses of the Thorndike position, but rather a somewhat confused philosophical-psychological broadside that strikes no mortal blow to the theory of identical elements.

HERBERT A. CARROLL.
Teachers College, Columbia University.

General Psychology for College Students, by Carl Newton Rexroad. New York: The Macmillan Co., 1929. Pp. XV + 392.

This text was written as an introduction to the behavioristic interpretation of psychology. It consequently contains a large amount of philosophical material. There are twenty-seven pages on science in general, and twenty more on the postulates of psychology as conceived by the author. The viewpoint characteristic of behavior-

ism is maintained. Living and non-living things are held to differ only in degree, and there is an interesting attempt to indicate a possible chemical explanation of life processes. Attention is compared to the selectivity of a radio. In the final chapter the author invades the field of ethics, defining the *summum bonum* as variability. In this, as in other respects, the view is that of Weiss.

The omissions and condensations of the book are, however, rather startling, and it seems to contain very little of the concrete experimental material found in most present day texts. The subjects of sensation and perception, presented in a chapter entitled "Man's Non-instrumental Analysis of the World," suffer most. Of the sense organs, only the eye is described, and the section on taste is just six lines long!

The chief excellence of the book is, in the opinion of the reviewer, the admirable summary of experiments on the conditioned reflex together with the diagrams illustrating the principles. There is also a good simple treatment of individual differences. The book is written in an interesting style and is easily read throughout, although some of the extreme condensations may be hardly intelligible to the beginning student. However, for instructors desiring to develop a behavioristic point of view as the introduction to other courses in the subject, the book will serve as an excellent instrument.

MELVIN RIGG.

Kenyon College.

Testing Intelligence and Achievement, by Albert J. Levine and Louis Marks. New York: The Macmillan Co., 1928. Pp. VIII + 399.

Mental measurement continues to be a favorite topic with educators. The increased application of measurement techniques has brought forth new discussions of their meaning and value. Among these is the present volume prepared by authorities of the New York City Schools as a textbook for introductory work in measurements. Their treatment is more comprehensive than the title would indicate for it includes such topics as Gestalt Psychology, The Learning Process, The Ductless Glands, The Causes of Psychopathy, The Education of the Gifted. Apparently the authors have attempted to combine a brief course in elementary psychology with a course in measurements, though the interrelation of the two fields is not always clear.

The topics most closely related to testing include the history of the measurement movement, the nature of intelligence and methods of measuring it, a description of representative individual tests and scales, of achievement tests, and of studies and tests of personality traits. With reference to the nature of intelligence the opinions of leading authorities are quoted. Methods of scale construction and the qualities of good tests and scales are described. Distinctive features and defects of the Stanford-Binet Scale, the Kuhlmann Revision and the Burt Scale are given. A list of representative achievement group tests is classified and annotated.

The authors observe that the study of personality traits other than intelligence has been greatly neglected in the past because of the difficulty of measuring emotion and character, and because of the popular feeling that these traits are not susceptible to measurement. In view of discrepancies in the intelligence and achievement of school children, it is their opinion that other measures of the child's personality are needed for a more complete understanding of the child. Attempts at measuring character traits and studies of personality are summarized in convenient form. The characteristics of mental defectives, of neurotics, and of gifted children are described together with methods of detecting deviates and suggestions for their school training. Reports of studies of gifted children are cited.

The organization of the material is not always logical and disgressions from the main topic of discussion frequently result in confusion. The evaluation of measurements is weakened somewhat by reservations and apologies which come long after we have been led to accept measurements as indispensable. The impression is gained that, after all, tests may be of little value and even dangerous. There is nowhere a scholarly treatment of the place of measurements, their specific uses and limitations. There is some misconception as to the relation of statistics and psychology. The authors remark that "It is somewhat difficult to tell where statistics end and psychology begins. Both are measures." Other statements are confusing or fantastic. What is meant by "in the subliminal depths of consciousness, below the level of the conscious level, below the level of the memory level . . .", "a too rigidly quantitative interpretation of a mentalistic personality, . . . " "the manysidedness of personality, the social, the psychological, the psychiatric . . . "? Confused sentence structure results in further vagaries. The careful student must disagree with the unqualified pronouncement that "Individual

tests are superior to group tests", and he will wonder why "The IQ derived from group tests should not be accorded the same consideration bestowed upon the Stanford IQ". Complete references to studies and names cited in the text are almost wholly lacking. Each chapter, however, contains a selected bibliography. There are examination questions covering each main topic and a summary of statistical symbols and formulæ needed for elementary work in measurements.

GERTRUDE HILDRETH.

The Lincoln School of Teachers College.